COVERT COMMUNICATION: FROM CLASSICAL CHANNELS TO QUANTUM CHANNELS

A Dissertation
Presented to
The Academic Faculty

By

Mehrdad Tahmasbi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

May 2020

COVERT COMMUNICATION:
FROM CLASSICAL CHANNELS TO QUANTUM CHANNELS

Approved by:

Dr. Matthieu Bloch, Advisor
School of Electrical Engineering
Georgia Institute of Technology

Dr. John Barry, Committee Chair
School of Electrical Engineering
Georgia Institute of Technology

Dr. Justin Romberg
School of Electrical Engineering
Georgia Institute of Technology

Dr. Brian Kennedy
School of Physics
Georgia Institute of Technology

Dr. Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Date Approved: March 24, 2020

ACKNOWLEDGEMENTS

I would like to first express my sincere gratitude toward my advisor Prof. Bloch. It is beyond words to describe how supportive he has been and how much I have learned from him. I also feel lucky to have had the chance to collaborate with and learn from Prof. Yener, Prof. Tan, Prof. Guha, and Prof. Bash in my research. I would like to thank Prof. Romberg, Prof. Barry, Prof. Koltchinskii, and Prof. Kennedy for kindly accepting to be on my PhD defense committee. It was also a great opportunity to work with excellent labmates: Keerthi, Ishaque, Meng-Che, Shi-Yuan, Nathan.

My graduate studies would have been difficult without the company of wonderful friends Vahid, Mojtaba, Mohammad, Ashkan, Mohammad Sadegh, Amirhossein, Homayoon, Majid, Ahmad, and Hossein.

This dissertation is dedicated to my family: Mom, Dad, and Milad.

TABLE OF CONTENTS

LIST OF FIGURES

NOTATION AND ABBREVIATIONS

Notation

$\mathbb{1}\{\cdot\}$	indicator function
$\log(\cdot)$	natural logarithm function
$\mathcal{X}, \mathcal{Y}, \cdots$	abstract sets
$\mathcal{X} \times \mathcal{Y}$	product of sets
$\mathcal{X}^n$	n-fold product of set $\mathcal{X}$
$\mathbf{x}, \mathbf{y}, \cdots$	sequences of length n
$\mathrm{wt}(\cdot)$	weight of binary sequences
$[\![a, b]\!]$	set of integers less than or equal to b and greater or equal to a
$O(\cdot), o(\cdot), \Omega(\cdot), \omega(\cdot), \Theta(\cdot)$	asymptotic Bachmann–Landau notations
$\left\|\cdot\right\|_p$	ℓ^p norm of vectors or Schatten norm of matrices
$X, Y, \cdots$	random variables
P_X	probability distribution of random variable X
P_X^{unif}	uniform probability distribution over $\mathcal{X}$
$\mathbb{E}_{P_X}(X)$	expected value of random variable X distributed according to P_X
$\mathrm{Var}_{P_X}(X)$	variance of random variable X distributed according to P_X
$P \otimes Q$	product distribution
$P^{\otimes n}$	n-fold product distribution
$P_X \otimes P_{Y\|X}$	joint distribution of (X, Y) associated with P_X and $P_{Y\|X}$
$P_{Y\|X} \circ P_X$	distribution of Y when (X, Y) distributed according to $P_X \otimes P_{Y\|X}$
$\mathbb{H}_b(\cdot)$	binary entropy function
$\mathbb{H}(X)$	entropy of random variable X
$H(P_X)$	entropy of probability distribution P_X
$\mathbb{I}(X;Y)$	mutual information between random variables X and Y
$I(P_X, P_{Y\|X})$	$\mathbb{I}(X;Y)$ where (X, Y) is distributed according to $P_X \times P_{Y\|X}$
$\mathbb{D}(\cdot\|\cdot)$	relative entropy

$\chi_2(\cdot,\cdot)$	chi-square divergence
$\beta_\alpha(P,Q)$	$\inf_{\mathcal{T}:P(\mathcal{T})\leqslant\alpha}(1-Q(\mathcal{T}))$ for two distributions P and Q
$\mathbb{Q}(\cdot)$	tail distributing function of standard normal distribution
$\dim\mathcal{H}$	dimensional of Hilbert space $\mathcal{H}$
$\|\phi\rangle, \|\psi\rangle, \cdots$	vectors in a Hilbert space over complex numbers
$\langle\phi\|$	linear functional associated with $\|\phi\rangle$
$\mathbf{1}_A$	identity operator on Hilbert space $\mathcal{H}_A$
$\mathcal{L}(\mathcal{H})$	set of all (bounded) linear operators on Hilbert space $\mathcal{H}$
$\mathcal{D}(\mathcal{H})$	set of density operators on Hilbert space $\mathcal{H}$
ρ_A^{unif}	mixed density operator on $\mathcal{H}_A$
id_A	identity operator on $\mathcal{L}(\mathcal{H}_A)$
$\mathrm{tr}\,(\cdot)$	trace of an operator
$\mathrm{tr}_A\,(\cdot)$	partial trace of a bipartite operator
$\nu(\cdot)$	number of distinct eigenvalues of an operator
$\lambda_{\min}(\cdot)$	minimum non-zero eigenvalue of a Hermitian operator
$X\succeq 0$	X is a positve semidefinite operator
$\mathcal{H}_A\otimes\mathcal{H}_B$	tensor product of two Hilbert spaces
$X\otimes Y$	tensor product of two operators

Abbreviation

AVC	Arbitrarly Varying Channel
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase-Shift Keying
BSC	Binary Symmetric Channel
CDF	Cumulative Distribution Function
CSI	Channel State Information
cq	classical-quantum
DMC	Discrete Memoryless Channel
iid	independent and identically distributed
IoT	Internet of Things
LPD	Low Probability of Detection
MAC	Multiple Access Channel
MLC	Multi-Level Coding
OOK	On–Off Keying
PDF	Probability Density Function
PMF	Probability Mass Function
POVM	Positive Operator Valued Measure
PPM	Pulse Position Modulation
QKD	Quantum Key Distribution
ROC	Receiver Operating Characteristic

SUMMARY

The objective of this dissertation is to study *covert communications* over classical and quantum channels. In contrast to the well-studied notion of secrecy, in which one attempts to protect the content of information, covertness (or low probability of detection) requires that the communication remains undetectable from an unwanted party, called the *warden*. This goal is achieved by using coding schemes that mimic the output statistics of the channel when there is no communication. A fundamental result states that the optimal number of transmitted bits scales as the square root of the number of channel uses when covertness is achieved. Many standard information-theoretic tools, therefore, fail in this zero-rate regime and one has to resort to the finite-length analysis of a protocol. In the first half of this dissertation, we establish results pertaining to the fundamental limits of covert communication over classical channels. In particular, we refine asymptotics of covert communications under three different metrics to measure covertness: relative entropy and variational distance between the probability distributions induced without and with communication, as well as the optimal probability of missed detection at a fixed probability of false alarm; we characterize the optimal throughput of covert communication over a non-coherent wireless channel; we study the effect of warden's knowledge of the code in covert communication. In the second half of this dissertation, we investigate the fundamental limits of covert secret key generation, in which two parties attempt to generate a secret key by using a classical or quantum channel and a public authenticated channel. We develop a scheme when the warden can arbitrarily change the state of a classical channel; we provide achievability results for covert secret key expansion over classical-quantum channels; we propose a scheme that could achieve positive throughput for covert secret key generation over a bosonic channel. Finally, in the last chapter, we study quantum steganography, in which the objective is to hide quantum information processing tasks when the warden has an inaccurate description of the channel.

CHAPTER 1
INTRODUCTION AND BACKGROUND

The concept of *information* plays a canonical role at the heart of several sciences and technologies. Analogous to Issac Newton who provided precise meanings well-suited for mathematical theories for the terms colloquially used in everyday conversations such as force, mass, etc., Claude Shannon developed the theory of information [1] in his 1948 paper *A mathematical theory of communication* [2]. He defined information as a sequence of symbols generated according to a random process and, in his words, considered the problem of "*reproducing at one point either exactly or approximately a message selected at another point*." One year later, Shannon published another paper to investigate the naturally related problem of information secrecy in his established framework of information theory. His formulation assumed no limit on the adversary's computational power but considered a statistical model for the message and the adversary's observations.

While information security concerns including, but not limited to, confidentiality, integrity, and authenticity of information and the methods by which they are alleviated can be traced back to ancient times, the development and dominance of computers with extremely high computational power and wireless networks in the last century has fundamentally altered the way we should think about these problems. Additionally, the surge of machine-to-machine wireless communications as part of the development of the IoT has escalated the privacy concerns and the risk of having sensitive data leaked. Although the celebrated RSA encryption/decryption algorithm is a successful solution in contemporary communication systems and has numerous desirable features, it also possesses unsatisfactory aspects, such as relying on limits on the computational power of the adversary. The perhaps complementary paradigm of information-theoretic security mentioned above could solve this issue. In this chapter, we describe two security problems that could be investigated

from an information-theoretic point of view, covert communication and secret key generation. We then review the quantum information theory formalism and discuss its relation to the secret key generation problem.

1.1 Covert Communication

1.1.1 Problem Formulation

While most information-theoretic security works to date have revolved around the issues of confidentiality and authentication [3, 4, 5], the growing concern around mass communication surveillance programs has reignited interest in investigating the covertness of communications, also known as LPD. In LPD problems, the objective is to hide the presence of communication and not necessarily to prevent information leakage about the messages transmitted. Following the analysis of LPD with space-time codes [6], recent works have investigated the information-theoretic limits of covert communications over noisy channel [7, 8]. In particular, building upon concepts from steganography [9], the study in [7] shows the existence of a "square-root law" for covert communication, which essentially states that no more than $O(\sqrt{n})$ bits can be communicated covertly over n channel uses of a memoryless channel.

We now review the formal model for point-to-point covert communication. As illustrated in Fig. 1.1, a legitimate transmitter (Alice) communicates with a legitimate receiver (Bob) over a memory-less channel $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ in the presence of a warden (Willie) who observes the communication through another memory-less channel $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$. There is no loss of generality in ignoring the joint channel $W_{YZ|X}$, as we shall see that results only depend on the marginal channels $W_{Y|X}$ and $W_{Z|X}$. We use shorthands $P_x \triangleq W_{Y|X=x}$, $P^{\mathbf{x}} \triangleq P_{x_1} \otimes \cdots \otimes P_{x_n}$, $Q_x \triangleq W_{Z|X=x}$, and $Q^{\mathbf{x}} \triangleq Q_{x_1} \otimes \cdots \otimes Q_{x_n}$. All parties are synchronized and the communication is supposed to happen over n channel uses. An "idle" symbol $0 \in \mathcal{X}$ corresponds to the expected input to the channel if no communication takes place, i.e., Willie's observation of the channel over n channel uses is distributed according to $Q^{\mathbf{0}}$

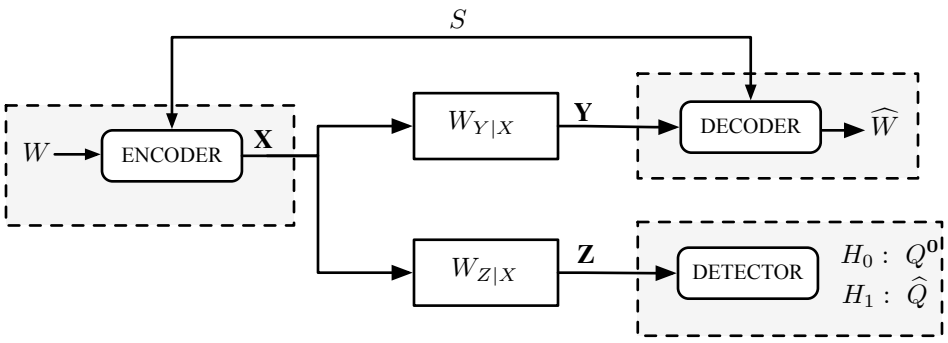

Figure 1.1: Model of covert communications over noisy channels

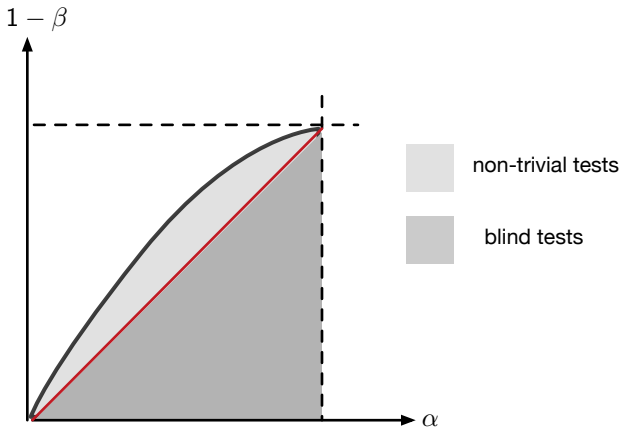

Figure 1.2: Performance of warden's tests

when there is no communication. The objective is for Alice to communicate a uniformly distributed message $W \in \mathcal{M}$ to Bob with a small probability of error while ensuring a low probability of detection from Willie. The communication may be assisted by a uniformly distributed shared secret key $S \in \mathcal{K}$. To this end, Alice encodes (W, S) through an encoder $f : \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n$ and Bob decodes W through a decoder $\phi : \mathcal{K} \times \mathcal{Y}^n \to \mathcal{M}$. We assume that the encoder and the decoder are known to all parties. Willie, on the other hand, uses its observations to detect the communication. Formally, the warden applies a possibly stochastic test $t : \mathcal{Z}^n \to [0, 1]$ to the output $\mathbf{z}$ of his channel and declares the existence of a communication with probability $t(\mathbf{z})$. Upon denoting the distribution of Willie's observations that is induced by the encoder f by $\widehat{Q} \triangleq \frac{1}{|\mathcal{M}||\mathcal{K}|} \sum_{m \in \mathcal{M}, k \in \mathcal{K}} Q^{f(m,k)}$, the false alarm

probability and the missed detection probability of the test t are

$$\alpha(t) \triangleq \mathbb{E}_{Q^{\mathbf{0}}}(t(\mathbf{Z})) \text{ and } \beta(t;f) \triangleq 1 - \mathbb{E}_{\widehat{Q}}(t(\mathbf{Z})), \tag{1.1}$$

respectively. Given an encoder f, we can illustrate Willie's performance by plotting the region $\mathcal{D} \triangleq \{(\alpha(t), 1-\beta(t;f)) : t : \mathcal{Z}^n \to [0,1]\}$ in a two-dimensional plane (see Fig. 1.2). Considering "blind" tests, which do not take the observations from the channel into account and are defined as $t(\mathbf{z}) \triangleq p$ for all $\mathbf{z}$ and for some $p \in [0,1]$, the set

$$\mathcal{B} \triangleq \{(\alpha, 1-\beta) : \alpha \geqslant 0, \beta \geqslant 0, \alpha + \beta \geqslant 1\}, \tag{1.2}$$

which is the region below the red line in Fig. 1.2, is always included in $\mathcal{D}$. This observation suggests the following *informal* criterion for covertness.

To ensure that Willie's observations through his channel do not help him detect the communication, the encoder f should be chosen such that $\mathcal{D} \approx \mathcal{B}$.

There is no single way of formalizing the above criterion as there exists several tests, and for each test t, there are two associated probabilities of error $\alpha(t)$ and $\beta(t;f)$. One commonly used quantity to measure covertness in the literature is the relative entropy between $\widehat{Q}$ and $Q^{\mathbf{0}}$, i.e., $\mathbb{D}\Big(\widehat{Q} \| Q^{\mathbf{0}}\Big)$, for the following two reasons. First, by Pinsker's inequality, we have

$$\alpha(t) + \beta(t;f) \geqslant 1 - \sqrt{\mathbb{D}\Big(\widehat{Q} \| Q^{\mathbf{0}}\Big)}\ \forall t, \tag{1.3}$$

which implies that by guaranteeing that $\mathbb{D}\Big(\widehat{Q} \| Q^{\mathbf{0}}\Big)$ is small, the performance is close to a blind test for any test t. In addition, one could take advantage of the relations of the relative entropy to other entropic quantities that appear in an information-theoretic analysis. We adopt this covertness measure throughout this chapter and we shall investigate other measures in the next chapters.

Remark 1. *The covert communication problem resembles steganography problem [10] in many aspects. In both problems, Willie's goal is to detect the existence of the communication, and similar scaling of the optimal number of bits has been established for both problems, known as the "square root law" [11]. However, the main difference between the two models is that the specific realization of the noise of the channel is unknown to all parties in covert communication, while the transmitter often exactly knows the cover text in steganography.*

1.1.2 The Fundamental Limits of Covert Communication: Square-Root-Law

We state here the main result for covert communication, known as square-root-law, which characterizes the asymptotically optimal number of bits that could be reliably and covertly transmitted. This result serves as the counter-part to the traditional Shannon channel coding theorem for covert communication.

Theorem 1. *Let $P_x \ll P_0$ and $Q_x \ll Q_0$ for all $x \in \mathcal{X}$ and $W_{Z|X} \circ P \neq Q_0$ for all $P \in \mathcal{P}(\mathcal{X})$ with $P(0) \neq 1$. Let $\mathcal{C}_n$ be a sequence of codes such that*

1. *$\mathcal{C}_n$ transmits a message of size M_n over n channel uses;*
2. *the average probability of error at the decoder vanishes as n gets large;*
3. *$\limsup_{n\to\infty} \mathbb{D}\Big(\widehat{Q}_n \| Q^{\mathbf{0}}\Big) \leqslant \delta$ where $\widehat{Q}_n$ is the distribution of Willie's observations induced by $\mathcal{C}_n$.*

The optimal scaling of $\log M_n$ *is* $\Theta(\sqrt{n})$*. Additionally, we have*

$$\sup_{\{\mathcal{C}_n\}_{n\geqslant 1}:(1),(2),(3)} \liminf_{n\to\infty} \frac{\log M_n}{\sqrt{n}} = \sup_{\widetilde{P}\in\mathcal{P}(\mathcal{X}):\widetilde{P}(0)=0} \sqrt{\frac{2\delta}{\chi_2(W_{Z|X}\circ\widetilde{P}\|Q_0)}} \sum_x \widetilde{P}(x)\mathbb{D}(P_x\|P_0). \tag{1.4}$$

The main implication of Theorem 1 is that if we normalize the number of transmitted bits by the number of channel uses, we asymptotically obtain zero. Instead, we normalize

the number of bits by the square root of the number of the channel uses to obtain a "covert throughput." The maximum achievable throughput is given in (1.4).

Sketch of Achievability Proof of Theorem 1

We sketch the achievability proof in three steps. The main ideas are based on [12, 13].

A generic protocol Analogous to many information-theoretic achievability proofs, we use a random codebook whose codewords are generated independently according to an iid distribution $P_n^{\otimes n}$ over $\mathcal{X}^n$. We shall specify later the distribution P_n, which depends on n, unlike standard random coding. We also suppose that the random codebook is instantiated from common randomness shared between Alice and Bob and unknown to the warden[1] to induce the distribution $W_{Z|X}^{\otimes n} \circ P_n^{\otimes n}$ at Willie's end. We address the de-randomization of the protocol in the last step of the proof. Let $\mathcal{M}_n$ be the message set and $\{\mathbf{X}_m \in \mathcal{X}^n\}_{m\in\mathcal{M}_n}$ be the random codebook. Alice transmits $\mathbf{X}_m$ when the message value is m. Note that the standard typical set decoding fails to achieve positive covert throughput because there exists a positive but arbitrary small rate penalty in that analysis, which asymptotically dominates the $O(\frac{1}{\sqrt{n}})$ rate of covert communication. We resort to a finite-length bound on the probability of error at the decoder (see Appendix 1.A)

$$\inf_{\gamma}\left[\mathbb{P}_{(P_n\otimes W_{Y|X})^{\otimes n}}\left(\log\frac{W_{Y|X}^{\otimes n}(\mathbf{Y}|\mathbf{X})}{W_{Y|X}^{\otimes n}\circ P_n^{\otimes n}(\mathbf{Y})}\leqslant\gamma\right)+\frac{|\mathcal{M}_n|}{\gamma}\right], \tag{1.5}$$

and choose P_n and $|\mathcal{M}_n|$ so that the above quantity vanishes.

Covert Process We now discuss how the distribution P_n should be chosen. First note that if P_n does not depend n, $\mathbb{D}\Big(W_{Z|X}^{\otimes n}\circ P_n^{\otimes n}\|Q_0^{\otimes n}\Big) = n\mathbb{D}\big(W_{Z|X}\circ P_n\|Q_0\big)$ diverges (or intuitively, by increasing the number of samples the warden can improve its test). Therefore, we have to ensure that by increasing n, $W_{Z|X}\circ P_n$ gets closer to Q_0. Let $\widetilde{P}$ be a distribution

[1]This is a non-uniform shared key, inconsistent with our assumption in Section 1.1.1. We shall, however "uniformize" the shared key in the last step of the proof.

over $\mathcal{X}$ such that $\widetilde{P}(0) = 0$ and define P_n as

$$P_n(x) \triangleq \begin{cases} 1 - \sqrt{\frac{2\delta}{\chi_2(W_{Z|X} \circ \widetilde{P} \| Q_0)}} n^{-\frac{1}{2}} & x = 0, \\ \sqrt{\frac{2\delta}{\chi_2(W_{Z|X} \circ \widetilde{P} \| Q_0)}} n^{-\frac{1}{2}} \widetilde{P}(x) & x \neq 0. \end{cases} \tag{1.6}$$

Using Chebyshev's inequality, one can show that [12, Eq. (16)]

$$\lim_{n \to \infty} \mathbb{P}_{(P_n \otimes W_{Y|X})^{\otimes n}} \left(\log \frac{W_{Y|X}^{\otimes n}(\mathbf{Y}|\mathbf{X})}{W_{Y|X}^{\otimes n} \circ P_n^{\otimes n}(\mathbf{Y})} \leqslant n I(P_n, W_{Y|X}) - n^{\frac{1}{4}} \sqrt{\log n} \right) = 0. \tag{1.7}$$

Therefore, choosing $\log M_n = n I(P_n, W_{Y|X}) - n^{\frac{1}{4}} \log n$ the probability of error vanishes by (1.5). Furthermore, the Taylor expansion of $\mathbb{D}\big((W_{Z|X} \circ P_n)^{\otimes n} \| Q_0^{\otimes n}\big)$ and $I(P_n, W_{Y|X})$ with respect to $1 - P_n(0)$ yields [12, Eq. (34) and Eq. (36)] that

$$\lim_{n \to \infty} \sqrt{n} I(P_n, W_{Y|X}) = \sqrt{\frac{2\delta}{\chi_2(W_{Z|X} \circ \widetilde{P} \| Q_0)}} \sum_x \widetilde{P}(x) \mathbb{D}(P_x \| P_0), \tag{1.8}$$

$$\lim_{n \to \infty} \mathbb{D}\Big(W_{Z|X}^{\otimes n} \circ P_n^{\otimes n} \| Q_0^{\otimes n} \Big) = \delta. \tag{1.9}$$

We thus achieve the desired throughput and covertness.

De-randomization The amount of common randomness required in the protocol is much larger than the size of the transmitted message. This could be mitigated by a technique known as channel resolvability [14]. Let $C_1, \cdots, C_L$ be L independent instantiation of the random codebook, which are sampled before the transmission and are known to Willie. Alice and Bob share a uniformly distributed secret key of size L according to which they choose one of the codes $C_1, \cdots, C_L$. Conditioned on using the codebook C_ℓ, let $\epsilon(\ell)$ and $\widehat{Q}(\ell)$ be the probability of error at Bob and the induced output distribution at Willie, respectively. By the law of total probability, $\mathbb{E}_{C_1, \cdots, C_L}\left(\frac{1}{L} \sum_{\ell=1}^L \epsilon(\ell) \right)$ and $\mathbb{E}_{C_1, \cdots, C_L}\left(\frac{1}{L} \sum_{\ell=1}^L \widehat{Q}(\ell) \right)$ are the probability of the generic protocol and $W_{Z|X}^{\otimes n} \circ P_n^{\otimes n}$, respectively. Since $C_1, \cdots, C_\ell$ are independent $\frac{1}{L} \sum_{\ell=1}^L \epsilon(\ell)$ and $\frac{1}{L} \sum_{\ell=1}^L \widehat{Q}(\ell)$ would be close to their mean with high

probability when L is large. A careful analysis [13] shows that $\log L$ should scale as $O(\sqrt{n})$ so that

$$\frac{1}{L}\sum_{\ell=1}^{L}\epsilon(\ell) \to \mathbb{E}_{C_1,\cdots,C_L}\left(\frac{1}{L}\sum_{\ell=1}^{L}\epsilon(\ell)\right), \tag{1.10}$$

$$\frac{1}{L}\sum_{\ell=1}^{L}\widehat{Q}(\ell) \xrightarrow{\ell_1} \mathbb{E}_{C_1,\cdots,C_L}\left(\frac{1}{L}\sum_{\ell=1}^{L}\widehat{Q}(\ell)\right), \tag{1.11}$$

in probability.

Sketch of Converse Proof of Theorem 1

To establish a converse for covert communication, we first use Fano's inequality to bound the number of bits in terms of $I(\overline{P}_n, W_{Y|X})$ where $\overline{P}_n$ is the average input distribution induced by the code. We then invoke the covertness constraint to restrict the average induced input distribution. Let $\{\mathcal{C}_n\}_{n\geqslant 1}$ be a sequence of codes such that the probability of error at the decoder for $\mathcal{C}_n$ is ϵ_n and $\mathbb{D}\Big(\widehat{Q}_n \| Q^{\mathbf{0}}\Big) \leqslant \delta_n$. Let $W, S, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be the random variables corresponding to the message, the key, the input sequence, Bob's output, and Willie's output, respectively when $\mathcal{C}_n$ is used. Following the calculations in [13, Eq. (82)-(87)], we have

$$\log M_n \leqslant \frac{nI(\overline{P}_n, W_{Y|X}) + \mathbb{H}_b\left(\epsilon_n\right)}{1-\epsilon_n} \tag{1.12}$$

$$= \frac{n\mathbb{D}\big(\overline{P}_n \otimes W_{Y|X} \| \overline{P}_n \otimes (W_{Y|X} \circ \overline{P}_n)\big) + \mathbb{H}_b\left(\epsilon_n\right)}{1-\epsilon_n} \tag{1.13}$$

$$\leqslant \frac{n\mathbb{D}\big(\overline{P}_n \otimes W_{Y|X} \| \overline{P}_n \otimes P_0\big) + \mathbb{H}_b\left(\epsilon_n\right)}{1-\epsilon_n} \tag{1.14}$$

$$= \frac{n(1-\overline{P}_n(0))\sum_{x\neq 0}\widetilde{P}_n(x)\mathbb{D}(P_x\|P_0) + \mathbb{H}_b\left(\epsilon_n\right)}{1-\epsilon_n} \tag{1.15}$$

where $\overline{P}_n$ is the average of the distributions of $X_1, \cdots, X_n$ and

$$\widetilde{P}_n(x) \triangleq \begin{cases} \frac{\overline{P}_n(x)}{1-\overline{P}_n(0)} & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{1.16}$$

We now exploit the covertness constraint to upper-bound $1 - \overline{P}_n(0)$. By [12, 13],

$$\mathbb{D}\big(W_{Z|X} \circ \overline{P}_n \| Q_0\big) \leqslant \frac{\delta}{n}. \tag{1.17}$$

Using the Taylor expansion of $\mathbb{D}\big(W_{Z|X} \circ \overline{P}_n \| Q_0\big)$ in $1 - \overline{P}_n(0)$, we have [12, Eq. (36)]

$$\mathbb{D}\big(W_{Z|X} \circ \overline{P}_n \| Q_0\big) = \frac{(1 - \overline{P}_n(0))^2}{2} \chi_2(W_{Z|X} \circ \widetilde{P}_n, Q_0) + O((1 - \overline{P}_n(0))^3) \tag{1.18}$$

Combining (1.15), (1.17), and (1.18) yields the converse.

1.1.3 Beyond Point-to-Point Covert Communication

There have been several studies in recent years attempting to generalize the model discussed so far in various directions and scrutinize the assumptions therein. We review here a subset of such studies.

1. Since communication systems typically include more than a single receiver and transmitter, studying covertness in a multi-user network is of interest. The covert capacity region of MACs and broadcast channels has been studied in [15] and [16, 17, 18], respectively. Perhaps surprisingly, the covertness constraint helps us simplify the capacity region in multi-user problems, as the sum-rate constraint is inactive for MACs [15], and time-division is optimal for a class of broadcast channels [17]. The routing of information in a wireless network with multiple relay nodes and multiple wardens has been also considered in [19].

2. The knowledge of the warden about the statistics of its channel could impact the

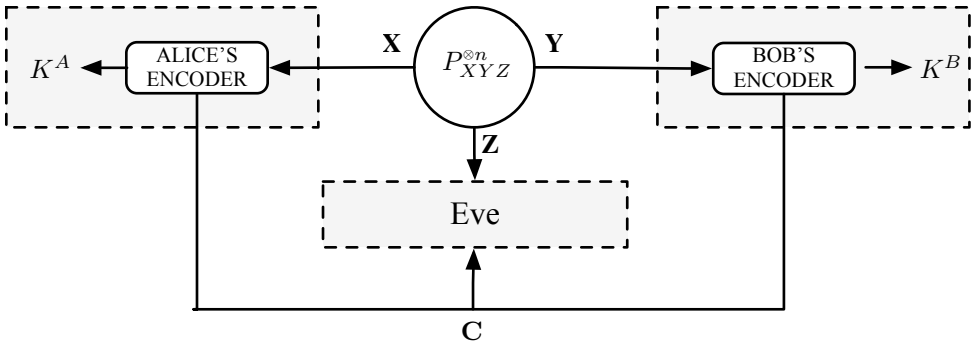

Figure 1.3: Model of secret key generation

scaling of the number of bits that could be transmitted covertly. In particular, the authors of [20] have shown that achieving a positive rate is possible if the warden has an inaccurate estimate of the channel noise.

3. The construction of efficient codes achieving positive covert throughput was studied in [21, 22]. The main challenge in the code construction is that no linear code would achieve the covert capacity [21, Lemma 1]. In [22], "a concatenated-style" code was proposed, while the coding scheme of [21] is based on PPM and MLC.

4. The authors of [23] have extended covert communication to classical-quantum channels. In [24], the authors considered quantum key distribution with a covertness constraint.

5. Covert communication when the warden does not know the exact transmission window has been studied in [25, 26], in which the scaling of the optimal number of bits changes to $O(\min(n, \sqrt{nT}))$, where T is the window size.

1.2 Secret Key Generation

To accomplish many cryptographic tasks, the legitimate parties (Alice and Bob) require common randomness unknown to unwanted parties (traditionally named Eve for eavesdropper), which we call a secret key. Legitimate parties can distill the secret key from correlated "partially" secure (in the sense that is formalized later) observations. To this

end, Alice and Bob should know the joint statistics of Eve with their observations and be connected by a bi-directional noiseless unlimited capacity channel whose output could be observed by Eve (see Fig. 1.3). Let $\mathbf{X} \in \mathcal{X}^n$, $\mathbf{Y} \in \mathcal{Y}^n$, and $\mathbf{Z} \in \mathcal{Z}^n$ denote the observations of Alice, Bob, and Eve, respectively, distributed according to $P_{XYZ}^{\otimes n}$. A secret key generation protocol operates as described next.

1. Alice and Bob interactively communicate over the public channel for r-rounds as follows. In the i^{th} round, Alice first transmits C_i^A, a function of $\mathbf{X}$, her local randomness, and previous communication over the public channel; Bob then transmits C_i^B, a function of $\mathbf{Y}$, his local randomness, and previous communication over the public channel.

2. After the r^{th} round, Alice and Bob use all their available information to distill keys K^A and K^B, respectively.

Let $\mathbf{C}$ denote all public communications and $P_{K^A K^B \mathbf{CZ}}$ denote the joint distribution of K^A, K^B, $\mathbf{Z}$, and $\mathbf{C}$ at the end of the protocol. We ask for two quantities $\mathbb{P}\big(K^A \neq K^B\big)$, $\big\|P_{K^A \mathbf{ZC}} - P_{K^A}^{\text{unif}} \otimes P_{\mathbf{ZC}}\big\|_1$ to be small. The first constraint ensures that the keys generated at Alice's and Bob's end are the same with high probability and the second constraint ensures that Eve has negligible knowledge about the generated key and the key is uniformly distributed.

Let us fix a probability distribution P_{XYZ} and a positive real number R. We say that the rate R is achievable if there exists a sequence of protocols $\{\mathcal{P}_n\}$ such that $\mathcal{P}_n$ generates keys of size 2^{nR} when $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ are distributed according to $P_{XYZ}^{\otimes n}$ and $\mathbb{P}\big(K^A \neq K^B\big)$ and $\big\|P_{K^A \mathbf{ZC}} - P_{K^A}^{\text{unif}} \otimes P_{\mathbf{ZC}}\big\|_1$ vanish in the limit of large n. In general, the maximum achievable rate for an arbitrary distribution P_{XYZ} is unknown, although many achievability and converse results have been established. The following theorem provides an achievability result.

Theorem 2. *The rate* $\max(0, \mathbb{I}(X;Y) - \min(\mathbb{I}(X;Z), \mathbb{I}(Y;Z)))$ *is achievable.*

Sketch of proof. It is enough to show that $\mathbb{I}(X;Y)-\mathbb{I}(Y;Z)$ is achievable when $\mathbb{I}(X;Y) > \mathbb{I}(Y;Z)$. We recall two coding techniques, information reconciliation and privacy amplification, essential to establish the achievability result.

Information Reconciliation Suppose Alice and Bob want to communicate over the public channel in order to agree on the same (possibly non-secure) piece of information. In particular, we prove in the following lemma how Bob can send a function of $\mathbf{Y}$ over the public channel so that Alice can decode $\mathbf{Y}$ with the help of $\mathbf{X}$.

Lemma 1 (Source coding with side information)**.** *Let $(\mathbf{X},\mathbf{Y})$ be distributed according to $P_{XY}^{\otimes n}$. Let $R > \mathbb{H}(Y|X)$ and $F : \mathcal{Y}^n \to [\![1, 2^{nR}]\!]$ be a random encoder that independently and uniformly at random maps $\mathbf{y}$ to an element of $[\![1, 2^{nR}]\!]$. There exists a decoder $\phi : \mathcal{X}^n \times [\![1, 2^{nR}]\!] \to \mathcal{Y}^n$ (depending on F) such that*

$$\lim_{n\to\infty} \mathbb{E}_F(\mathbb{P}(\phi(\mathbf{X}, F(\mathbf{Y})) \neq \mathbf{Y}|F)) = 0. \tag{1.19}$$

Sketch of proof. We adopt the standard notation of [27] for typical sets. Let ϕ be the typical set decoder, i.e., $\phi(\mathbf{x}, c) = \mathbf{y}$ if $\mathbf{y}$ is the unique element of $\mathcal{Y}^n$ such that $(\mathbf{x},\mathbf{y}) \in \mathcal{T}_n^\epsilon(X,Y)$ and $F(c) = \mathbf{y}$. Let $(\mathbf{x},\mathbf{y})$ be fixed. An error occurs if $(\mathbf{x},\mathbf{y})$ is not typical or there exists an element $\widetilde{\mathbf{y}} \in \mathcal{T}_n^\epsilon(Y|\mathbf{x})$ with $F(\mathbf{y}) = F(\widetilde{\mathbf{y}})$. The first probability goes to zero. To bound the second probability consider, note that $|\mathcal{T}_n^\epsilon(Y|\mathbf{x})| \approx 2^{n\mathbb{H}(Y|X)}$ and for each $\widetilde{\mathbf{y}}$, $\mathbb{P}(F(\widetilde{\mathbf{y}}) = F(\mathbf{y})) = 2^{-nR}$. Therefore, by the union bound the second probability goes to zero too. See [28] for a detailed proof. □

Privacy Amplification The objective here is to reduce Eve's information at the expense of decreasing the size of the initially shared information between Alice and Bob. The following lemma states that, if we randomly "bin" $\mathbf{Y}$, Eve obtains negligible information about the bin provided that the number of bins is not too large.

Lemma 2 ([29]). *Let* $(\mathbf{Y}, \mathbf{Z})$ *be distributed according to* $P_{YZ}^{\otimes n}$. *Let* $R < \mathbb{H}(Y|Z)$ *and* $F : \mathcal{Y}^n \to [\![1, 2^{nR}]\!]$ *be a random encoder that independently and uniformly at random maps* $\mathbf{y}$ *to an element of* $[\![1, 2^{nR}]\!]$. *We have*

$$\lim_{n\to\infty} \mathbb{E}_F\Big(\big\|P_{\mathbf{Z}F(\mathbf{Y})} - P^{\text{unif}}_{F(\mathbf{Y})} \otimes P_Z^{\otimes n}\big\|_1\Big) = 0. \tag{1.20}$$

We are now ready to prove the theorem. Let $F : \mathcal{Y}^n \to [\![1, 2^{nR_1}]\!] \times [\![1, 2^{nR_2}]\!]$ be a random encoder where each output is chosen independently at random and R_1 and R_2 satisfy $R_1 + R_2 < \mathbb{H}(Y|Z)$ and $R_2 > \mathbb{H}(Y|X)$. We can always choose such R_1 and R_2 with R_1 arbitrary close to $\mathbb{I}(X;Y) - \mathbb{I}(Y;Z)$. Defining $(C, K) = F(\mathbf{Y})$, Bob transmits C over the channel, Alice decodes $\mathbf{Y}$ from C and $\mathbf{X}$. Alice then reconstructs K from her estimate of $\mathbf{Y}$. By Lemma 1, Alice decodes the message with high probability and by Lemma 2, $P_{C,K,\mathbf{Z}}$ is close to $P^{\text{unif}}_{C,K} \otimes P_Z^{\otimes n}$. □

Remark 2. *The function* F *in both Lemma 1 and 2 could be chosen uniformly from a family of two-universal functions [30] from* $\mathcal{Y}^n$ *to* $[\![1, 2^{nR}]\!]$. *This could reduce the sample space of* F *significantly. Furthermore, although one can argue that there exists a specific instantiation* f *of* F *with good performance, Bob could always randomly sample* F *and transmit it over the public channel.*

1.3 Quantum Information Theory

1.3.1 Preliminaries

In quantum information theory, one revisits the Shannon theory of information by considering the quantum nature of information-processing devices. We review here the formalism of quantum mechanics (see e.g., [31, 32] for a complete introduction).

Associated with each system A, there exists a Hilbert space $\mathcal{H}_A$; the state of the system is described by a positive unit-trace linear operator $\rho_A : \mathcal{H}_A \to \mathcal{H}_A$. Such operator is called a density operator and $\mathcal{D}(\mathcal{H}_A)$ denotes the set of all density operators acting on $\mathcal{H}_A$. A state

is pure if it is of the form $|\phi\rangle\langle\phi|_A$ for some unit vector $|\phi\rangle_A$. We often use ϕ_A to denote $|\phi\rangle\langle\phi|_A$. Let $\mathbf{1}_A$ denote the identity operator acting on $\mathcal{H}_A$. A measurement is a set $\mathbf{M} = \{M_i\}_{i\in\mathcal{I}}$ such that $M_i \succeq 0$ for all $i \in \mathcal{I}$ and $\sum_{i\in\mathcal{I}} M_i = \mathbf{1}_A$. The output of the measurement $\mathbf{M}$ on a system in the state ρ_A would be $\text{tr}\,(\rho_A M_i)$. The Hilbert space corresponding to the composition of the systems A and B is $\mathcal{H}_A \otimes \mathcal{H}_B$. If the joint system is in the state ρ_{AB}, the sub-systems A and B are in the states $\rho_A = \text{tr}_B\,(\rho_{AB})$ and $\rho_B = \text{tr}_A\,(\rho_{AB})$, respectively. For a state ρ_A, we denote the von Neumann entropy by $H(\rho_A) \triangleq \mathbb{H}(A)_\rho = -\text{tr}\,(\rho_A \log \rho_A)$. For a joint state ρ_{AB}, von Neumann conditional entropy and mutual information are defined as $\mathbb{H}(A|B)_\rho \triangleq \mathbb{H}(AB)_\rho - \mathbb{H}(B)_\rho$ and $\mathbb{I}(A;B)_\rho \triangleq \mathbb{H}(A)_\rho - \mathbb{H}(A|B)_\rho$, respectively. For two states ρ and σ for the same system, we define three distances: the quantum relative entropy $\mathbb{D}(\rho\|\sigma) \triangleq \text{tr}\,(\rho(\log\rho - \log\sigma))$ (if $\text{support}\,(\sigma) \nsubseteq \text{support}\,(\rho)$, we define $\mathbb{D}(\rho\|\sigma) = \infty$), the trace distance $\|\sigma - \rho\|_1 \triangleq \text{tr}\left(\sqrt{(\sigma-\rho)^\dagger(\sigma-\rho)}\right)$, and the fidelity $F(\rho,\sigma) \triangleq \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2$. A quantum process with input system A and output system B can be formally described by a completely-positive trace-preserving map $\mathcal{N} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$. The identity channel mapping each state in $\mathcal{D}(\mathcal{H}_A)$ to itself shall be denoted by id_A. We recall here some important facts that will be frequently used in the last three chapters.

Data Processing Inequality Let ρ and σ be in $\mathcal{D}(\mathcal{H}_A)$ and $\mathcal{N} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$. We have

$$\mathbb{D}(\rho\|\sigma) \geqslant \mathbb{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \tag{1.21}$$

$$\|\rho - \sigma\|_1 \geqslant \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 \tag{1.22}$$

$$F(\rho,\sigma) \leqslant F(\mathcal{N}(\rho),\mathcal{N}(\sigma)) \tag{1.23}$$

Spectral Decomposition All density operators are of the form $\sum_i \lambda_i |e_i\rangle\langle e_i|$ where $\lambda_i \geqslant 0$, $\sum_i \lambda_i = 1$, and $\{|e_i\rangle\}$ is an orthonormal basis for the underlying Hilbert space.

Purification For any state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$, there exists another system, called reference, and a pure state $|\phi\rangle_{RA} \in \mathcal{H}_A \otimes \mathcal{H}_R$ such that $\rho_A = \mathrm{tr}_R(\phi_{RA})$.

Schmit Decomposition Let $|\phi\rangle_{AB}$ be a state in $\mathcal{H}_A \otimes \mathcal{H}_B$. There exists orthonormal family of vectors $\{|e_i\rangle\}_{i\in\mathcal{I}} \subset \mathcal{H}_A$ and $\{|f_i\rangle\}_{i\in\mathcal{I}} \subset \mathcal{H}_B$ and positive real numbers $\{\lambda_i\}_{i\in\mathcal{I}}$ with $\sum_i \lambda_i = 1$ such that $|\phi\rangle_{AB} = \sum_{i\in\mathcal{I}} \sqrt{\lambda_i}|e_i\rangle \otimes |f_i\rangle$. One important consequence of this results is that $\mathbb{H}(\phi)_A = \mathbb{H}(\phi)_B$ for any pure state $|\phi\rangle_{AB}$.

Uhlmann's Theorem Let σ and ρ be two density operators acting on the same space. Let $|\phi\rangle_{AR}$ and $|\psi\rangle_{AR}$ be purifications of ρ and σ, respective. We have [33]

$$F(\rho,\sigma) = \max_{U:\text{unitary on } R} |\langle\psi|(\mathbf{1}_A \otimes U)|\phi\rangle|^2. \tag{1.24}$$

1.3.2 Quantum Key Distribution

We now discuss an application of quantum information theory, called quantum key distribution. Note that in the secret key generation problem introduced in Section 1.2, a crucial assumption is the knowledge of the statistics of the eavesdropper's observation. The quantum mechanics laws allow Alice and Bob to circumvent this assumption and to upper-bound Eve's information using only the statistics of their data. In particular, let Alice, Bob and Eve share a tripartite state ρ_{ABE}. We claim that

$$\mathbb{I}(A;E)_\rho \leqslant \mathbb{H}(A)_\rho + \mathbb{H}(A|B)_\rho. \tag{1.25}$$

To see this, we take a purification $|\phi\rangle_{ABER}$ of ρ_{ABE} and note that

$$\mathbb{I}(A;E)_\rho \leqslant \mathbb{I}(A;ER)_\phi = \mathbb{H}(A)_\phi + \mathbb{H}(ER)_\phi - \mathbb{H}(AER)_\phi = \mathbb{H}(A)_\rho + \mathbb{H}(A|B)_\rho. \tag{1.26}$$

Since privacy amplification works even when Eve's observations are quantum [30], Alice and Bob can ensure the security of the generated key by knowing only the statistics of their observations. Note that to achieve a positive rate we need $\mathbb{I}(A;B)_\rho - \mathbb{H}(A)_\rho + \mathbb{H}(A|B)_\rho > 0$, which simplifies as $\mathbb{H}(A|B)_\rho < 0$. This inequality holds if and only if ρ_{AB} is entangled. Based on this idea, numerous protocols have been introduced (see the survey [34]) for quantum key distribution, overcoming many theoretical and experimental challenges.

1.4 Outline of the Dissertation and Related Publications

In the first three chapters of this dissertation, we consider the fundamental limits of covert communication. In the following three chapters, we investigate the problem of secret key generation over classical or quantum channels with covertness constraint. In the last chapter, we study a related problem to covert communication, called quantum steganography.

In **Chapter 2**, we consider the finite block-length performance for covert communication under three covertness metrics: relative entropy, total variation, and minimum missed detection at a fixed probability of false alarm. We find the optimal first-order asymptotics for binary-input DMC for all three metrics. We also characterize the optimal second-order asymptotics when the covertness is measured by the relative entropy and provide lower- and upper-bounds on the second-order when the covertness is measured through total variation, and minimum missed detection at a fixed probability of false alarm. The results of this chapter are based on the following publications.

- M. Tahmasbi and M. R. Bloch, "First and Second Order Asymptotics in Covert Communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

- M. Tahmasbi and M. R. Bloch, "Second-Order Asymptotics of Covert Communications over Noisy Channels," in Proc. of IEEE International Symposium on Information Theory, Barcelona, Spain, Jul. 2016, pp. 2224–2228.

In **Chapter 3**, we consider the problem of covert communication over non-coherent fast Rayleigh-fading wireless channel. This problem is motivated by the fact the covert capacity of AWGN channels is achieved by vanishing amplitude distributions, which makes the detection of the phase difficult at the receiver. We characterize the covert capacity as the solution to an optimization problem and show that the optimal distribution is discrete with a finite number of mass points and establish an amplitude constraint on the position of mass points. The results of this chapter are based on the following manuscript.

- M. Tahmasbi, A. Savard, and M. R. Bloch, "Covert Capacity of Non-Coherent Rayleigh-Fading Channels." submitted to *IEEE Transactions on Information Theory*, Oct. 2018.

In **Chapter 4**, we analyze how the knowledge of the warden about the code affects the ability of the legitimate users to communicate covertly. We first show that when the code is unknown, the square root law still holds. In addition, we consider the scenario in which the warden does not know the code but has previous samples of its channel when the code was used. We prove asymptotically tight lower- and upper-bounds on the number samples required to "learn" the code. The results of this chapter are based on the following manuscript.

- M. Tahmasbi and M. R. Bloch, "Covert communication when the code is unknown at the warden," accepted to Allerton Conference, August 2019.

In **Chapter 5**, we propose a formal model for secret and covert key generation using public communication and n instances of a noisy channel. We develop achievability and converse bounds for a passive warden. We then extend the achievability results for an active warden who can arbitrarily change the state of the channel. The results of this chapter are based on the following publications.

- M. Tahmasbi and M. R. Bloch, "Covert secret key generation," in Proc. of IEEE Conference on Communications and Network Security, Workshop on Physical-Layer

Methods for Wireless Security, Las Vegas, NV, Oct. 2017, pp. 540–544.

- M. Tahmasbi and M. R. Bloch, "Covert Secret Key Generation With an Active Warden," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1026–1039, 2020.

In **Chapter 6**, we extend the results of Chapter 5 for a cq channels. This requires a quantum channel resolvability result and a careful analysis of the error exponents. The results of this chapter are based on the following publication.

- M. Tahmasbi and M. R. Bloch, "Framework for covert and secret key expansion over classical-quantum channels," *Physical Review A*, vol. 99, no. 5, p. 052329, May 2019.

In **Chapter 7**, we develop a protocol for quantum key distribution running over bosonic channels satisfying a covertness constraint. Our protocol achieves the square root law with (very) low throughput. While our results are slightly disappointing in that the range of useful parameters is limited, they open the way to experimental demonstrations of covert QKD. The results of this chapter are based on the following manuscript.

- M. Tahmasbi and M. R. Bloch, "Towards undetectable quantum key distribution over bosonic channels," submitted to *Journal on Selected Areas in Information Theory*.

Finally, in **Chapter 8**, we establish results for quantum steganography, in which the warden has an inaccurate knowledge of the channel and the legitimate parties leverage this lack of knowledge to perform a quantum information processing task without disturbing the output statistics of the channel expected by the warden. We formalize the problem for four combinations of quantum information processioning tasks, for which we establish achievability results. The results of this chapter are based on the following references.

- M. Tahmasbi and M. Bloch, "Steganography Protocols for Quantum Channels," in Proc. of IEEE International Symposium on Information Theory, Paris, France, Jul. 2019, pp. 2179–2183

- M. Tahmasbi, M. R. Bloch, "Steganography Protocols for Quantum Channels," submitted to *Journal of Mathematical Physics*.

APPENDIX

1.A One-Shot Coding Results

As discussed in Section 1.1.2, we frequently resort to one-shot coding results in achievability proofs of covert communication. We gather here the most useful results for the reader's convenience.

1.A.1 Classical Channel Coding

As depicted in Fig. 1.A.1, let $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ be a classical DMC, W be a message uniformly distributed over $\mathcal{M}$, and $f : \mathcal{M} \to \mathcal{X}^n$ be an encoder. The decoder forms the maximum likelihood estimate of W as $\widehat{W} \triangleq \arg\max_{m\in\mathcal{M}} W_{Y|X}^{\otimes n}(\mathbf{Y}|f(m))$. We then have the following result in this setup.

Lemma 3 ([35])**.** *Let $P_{\mathbf{X}}$ be a probability distribution over $\mathcal{X}^n$ and $F : \mathcal{M} \to \mathcal{X}^n$ be a random encoder whose codewords are iid according to a distribution $P_{\mathbf{X}}$. Let $0 < \epsilon' < \epsilon < 1$ and*

$$\log|\mathcal{M}| \geqslant \mathbb{D}_s^{\epsilon-\epsilon'}\left(P_{\mathbf{X}} \otimes W_{Y|X}^{\otimes n} \| P_{\mathbf{X}} \otimes (W_{Y|X}^{\otimes n} \circ P_{\mathbf{X}})\right) + \log\epsilon', \tag{1.27}$$

where $\mathbb{D}_s^{\eta}(P\|Q) \triangleq \sup\left\{\gamma : \mathbb{P}_P\left(\log\frac{P(X)}{Q(X)} \leqslant \gamma\right) \leqslant \eta\right\}$. We then have $\mathbb{E}_F\left(\mathbb{P}\left(\widehat{W} \neq W\right)\right) \leqslant \epsilon$.

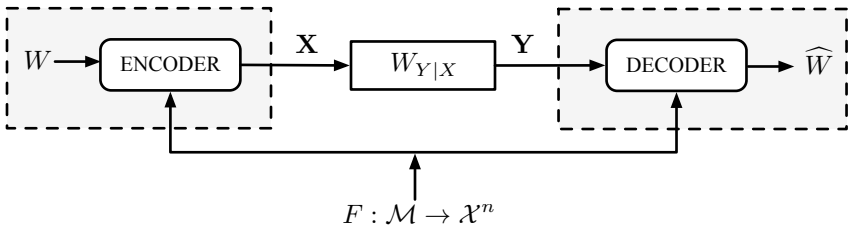

Figure 1.A.1: Random coding for reliable communication

We now replace $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ by a cq-channel $x \mapsto \rho_B^x$. Note that the decoder should be a quantum measurement $\{\Gamma_m\}_{m\in\mathcal{M}}$ while the encoder is still a classical function $f : \mathcal{M} \to \mathcal{X}^n$.

Lemma 4. ([36, Theorem 1]) *Let $P_{\mathbf{X}}$, $F : \mathcal{M} \to \mathcal{X}^n$, ϵ, ϵ' be as in Lemma 3. Let*

$$\log M \leqslant \mathbb{D}_H^{\epsilon-\epsilon'}(\rho_{\mathbf{XB}} \| \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) + \log \epsilon', \tag{1.28}$$

where $\rho_{\mathbf{XB}} \triangleq \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \rho_{\mathbf{B}}^{\mathbf{x}}$, $\rho_{\mathbf{B}}^{\mathbf{x}} \triangleq \rho_B^{x_1} \otimes \cdots \otimes \rho_B^{x_n}$, and

$$\mathbb{D}_H^{\eta}(\rho\|\sigma) \triangleq -\log \inf_{Q:0\leqslant Q\leqslant \mathbf{1}, \mathrm{tr}(Q\rho)\geqslant 1-\eta} \mathrm{tr}\left(Q\sigma\right). \tag{1.29}$$

We then have for a POVM $\{\Gamma_m\}_{m\in\mathcal{M}}$ that could possibly depend on F,

$$\mathbb{E}_F \left(\frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \mathrm{tr}\left(\Gamma_m \rho_{\mathbf{B}}^{F(m)}\right) \right) \geqslant 1 - \epsilon. \tag{1.30}$$

1.A.2 Classical and Quantum Channel Resolvability

Let $(\mathcal{X}, W_{Z|X}, \mathcal{Y})$ be a classical DMC and $Q^{\mathbf{x}}$ be defined as in Section 1.1.1. The following result holds, which is known as channel resolvability.

Lemma 5. *Let $P_{\mathbf{X}}$ be a probability distribution over $\mathcal{X}^n$ and $F : \mathcal{M} \to \mathcal{X}^n$ be a random encoder whose codewords are independently sampled from $P_{\mathbf{X}}$. Let $\mathcal{M}$, δ, δ', λ be such that*

$$\log |\mathcal{M}| \geqslant \mathbb{D}_s^{1-\delta+\delta'} \left(P_{\mathbf{X}} \otimes W_{Z|X}^{\otimes n} \| P_{\mathbf{X}} \otimes Q^{\mathbf{0}} \right) + \log \frac{1}{4\delta'^2} \tag{1.31}$$

We then have

$$\mathbb{E}_F \left(\frac{1}{2} \left\| \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} Q^{F(m)} - W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} \right\|_1 \right) \leqslant \delta. \tag{1.32}$$

and

$$\mathbb{P}_F\left(\frac{1}{2}\left\|\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} Q^{F(m)} - W_{Z|X}^{\otimes n}\circ P_{\mathbf{X}}\right\|_1 \geqslant \delta+\lambda\right) \leqslant \exp\left(-2\,|\mathcal{M}|\,\lambda^2\right). \tag{1.33}$$

Proof. (1.31) follows from [13, Lemma 2] and (1.33) follows from McDiarmid's inequality (see Theorem 5 in Appendix 1.B). □

Let us replace $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ by a cq-channel $x \mapsto \rho_W^x$. We have the following result.

Lemma 6. ([37, Lemma 9.2]) *Let $P_{\mathbf{X}}$ and F be as in Lemma 3. Let $\rho_{\mathbf{W}}^{\mathbf{x}} \triangleq \rho_W^{x_1}\otimes\cdots\otimes\rho_W^{x_n}$. We have*

$$\mathbb{E}_F\left(\left\|\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\rho_{\mathbf{W}}^{F(m)} - \rho_{\mathbf{W}}\right\|_1\right) \leqslant 2\sqrt{2^{\gamma s+\phi(s)}} + \sqrt{\frac{2^{\gamma}\nu(\rho_{\mathbf{W}})}{|\mathcal{M}|}}, \tag{1.34}$$

where $\phi(s) \triangleq \log\left(\sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x})\mathrm{tr}\left((\rho_{\mathbf{W}^{\mathbf{x}}}^{1-s}\rho_{\mathbf{W}}{}^{s}\right)\right)$.

1.A.3 Privacy Amplification with Quantum Side Information

We first recall the definition of a two-universal family of hash functions.

Definition 1. *Let $\mathcal{X}$ and $\mathcal{Z}$ be two finite non-empty sets. A non-empty family of functions $\mathcal{F}$ from $\mathcal{X}$ to $\mathcal{Z}$ is called* two-universal *if for all distinct $x, x' \in \mathcal{X}$, we have*

$$\frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}}\mathbb{1}\{f(x)=f(x')\} \leqslant \frac{1}{|\mathcal{Z}|}. \tag{1.35}$$

Moreover, $\mathcal{F}$ is called regular *if for all $f\in\mathcal{F}$ and all $z\in\mathcal{Z}$, we have $|f^{-1}(z)| = \frac{|\mathcal{X}|}{|\mathcal{Z}|}$, where $f^{-1}(z) \triangleq \{x\in\mathcal{X} : f(x)=z\}$.*

The next two results are well-known properties of two-universal hash functions.

Proposition 1. *Let $\mathcal{X}$ and $\mathcal{Z}$ be two non-empty finite sets such that $|\mathcal{X}|$ is divisible by $|\mathcal{Z}|$. There exists a two-universal regular family of functions from $\mathcal{X}$ to $\mathcal{Z}$.*

Proof. All functions f with $f^{-1}(z) = \frac{|\mathcal{Z}|}{|\mathcal{X}|}$ form a two-universal family of hash functions. □

Lemma 7. ([30]) *Let ρ_{XA} be a cq-state on $\mathcal{H}_X \otimes \mathcal{H}_A$ with respect to an orthonormal basis $\{|x\rangle : x \in \mathcal{X}\}$ for $\mathcal{H}_X$, and $\mathcal{F}$ be a two-universal family of functions from $\mathcal{X}$ to $\mathcal{Z}$. We then have*

$$\frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}}\left\|(\mathcal{E}^f_{X\to Z}\otimes \mathrm{id}_A)(\rho_{XA}) - \rho_Z^{\mathrm{unif}}\otimes\rho_A\right\|_1 \leqslant \inf_{\epsilon\geqslant 0}\left[2\epsilon + 2^{-\frac{1}{2}\left(\mathbb{H}^\epsilon_{\min}(X|A)_\rho - \log|\mathcal{Z}|\right)}\right], \tag{1.36}$$

where $\mathcal{E}^f_{X\to Z} : \rho_X \mapsto \sum_{x\in\mathcal{X}} |f(x)\rangle\langle x|\rho_X|x\rangle\langle f(x)|$.

1.B Concentration Inequalities

We gather here four results from concentration of measure theory, which could be applied to the one-shot results in Appendix 1.A when the channel is memory-less. A reader could refer to standard texts on high dimensional probabilities for details and proofs (e.g., [38]).

Theorem 3 (Hoeffding's Inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables such that $a_i \leqslant X_i \leqslant b_i$ almost surely for all $i \in [\![1, n]\!]$. We have for all $t > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{1.37}$$

Theorem 4 (Bernstein Inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables such that $|X_i - \mathbb{E}(X_i)| \leqslant M$ almost surely for all $i \in [\![1, n]\!]$. We have for all $t > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \geqslant t\right) \leqslant \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathrm{Var}(X_i) + \frac{1}{2}Mt}\right). \tag{1.38}$$

Theorem 5 (McDiarmid's Inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables*

belonging to a set $\mathcal{X}$ and $g : \mathcal{X}^n \to \mathbb{R}$ be a function satisfying

$$\sup_{x_1,\cdots,x_n,x_i'} |g(x_1,\cdots,x_i,\cdots,x_n) - g(x_1,\cdots,x_i',\cdots,x_n)| \leqslant c_i, \quad \forall i \in [\![1,n]\!]. \tag{1.39}$$

We have for all $t \geqslant 0$,

$$\mathbb{P}(g(\mathbf{X}) - \mathbb{E}(g(\mathbf{X})) \geqslant t) \leqslant \exp\left(\frac{-2t^2}{\sum c_i^2}\right). \tag{1.40}$$

Theorem 6 (Berry-Esseen Theorem)**.** *Let $X_1,\cdots,X_n$ be independent random variables such that for $k \in [\![1,n]\!]$ we have $\mathbb{E}(X_k) = \mu_k$, $\sigma_k^2 = \text{Var}(X_k)$, and $t_k = \mathbb{E}(|X_k - \mu_k|^3)$. If we define $\sigma^2 = \sum_{k=1}^n \sigma_k^2$ and $T = \sum_{k=1}^n t_k$, then we have*

$$\left|\mathbb{P}\left(\sum_{k=1}^n (X_k - \mu_k) \geqslant \lambda\sigma\right) - \mathbb{Q}(\lambda)\right| \leqslant \frac{6T}{\sigma^3}. \tag{1.41}$$

CHAPTER 2
FUNDAMENTAL LIMITS OF COVERT COMMUNICATION: SECOND-ORDER ASYMPTOTICS

2.1 Summary

We study the first- and second-order asymptotics of covert communication over binary-input DMCs for three different covertness metrics and under the maximum probability of error constraint. When covertness is measured in terms of the relative entropy between the channel output distributions induced with and without communication, we characterize the exact first- and second-order asymptotics of the number of bits that can be reliably transmitted with a maximum probability of error less than ϵ and a relative entropy less than δ. When covertness is measured in terms of the variational distance between the channel output distributions or in terms of the probability of missed detection for a fixed probability of false alarm, we establish the exact first-order asymptotics and bound the second-order asymptotics. PPM achieves the optimal first-order asymptotics for all three metrics, as well as the optimal second-order asymptotics for relative entropy. The main conceptual contribution of this paper is to clarify how the choice of a covertness metric impacts the information-theoretic limits of covert communications. The main technical contribution underlying our results is a detailed expurgation argument to show the existence of a code satisfying the reliability and covertness criteria. The content of this chapter is based on [39, 40].

2.2 Introduction

The square-root law of covert communication has been refined in several follow-up works [41]; in particular, the information-theoretic limits are now known for classical discrete and

Gaussian memoryless channels [13, 12], classical-quantum channels [42, 23], and multiple-access channels [43], when covertness is measured in terms of the relative entropy between the channel output distributions induced with and without communication. Note that the choice of relative entropy as a metric for covertness is guided in part by the natural connection between relative entropy and information-theoretic metrics, such as entropy and mutual information, which have been largely explored in the context of information-theoretic security [44].

The contribution of the present paper is twofold. First, as an attempt to develop operational characterizations of covertness, we study the information-theoretic limits of covert communication for alternative metrics, including variational distance, and the probability of missed detection. Second, motivated by the likely time-limited nature of covert communications, we make a first step towards a finite-length analysis and extend the first-order analysis of information-theoretic limits to second-order asymptotics. The specific results developed in the present paper focus on binary-input pDMC to yield simple closed-form expressions and are the following.

- We characterize the exact second-order asymptotics of the maximum number of reliable and covert bits that can be transmitted with maximum probability of error ϵ and *relative entropy* δ between channel output distributions with and without communication (Theorem 7); this corrects an unfortunate error in the conference version [45], in which we claimed erroneous second-order asymptotics for arbitrary codes with an average probability of error ϵ.

- We characterize the exact first-order and bound the second-order asymptotics of the maximum number of reliable and covert bits that can be transmitted with maximum probability of error ϵ and *variational distance* δ between channel output distributions with and without communication (Theorem 8).

- Finally, we characterize the exact first-order and bound the second-order asymptotics

of the maximum number of reliable and covert bits that can be transmitted with maximum probability of error ϵ and probability of missed detection $1-\alpha-\delta$ when α is the adversary's probability of false alarm (Theorem 9).

All our achievability results are established using PPM, which optimality was previously only established for the first-order asymptotics with relative entropy [46]. The operational relevance of codes used in conjunction with PPM, which may be viewed as a highly structured subset of constant-composition codes, is justified by recent work towards practical code design [21], in which PPM plays a crucial role. We also emphasize from the outset that the focus on the maximal probability of error is essential to our analysis. As discussed in Remark 3 and Appendix 2.A, there is no strong converse for average probability of error without additional constraint; hence, considering the *maximum* probability of error as the reliability metric is reasonable for second-order analysis.

The second-order asymptotics obtained with a relative entropy metric for covertness are what could have been expected by extrapolating the results of channel coding in the finite length regime [35] to the first-order asymptotics of covert communication [13, 12]; however, the proof requires specific techniques beyond those used to study the first-order asymptotics of covert communications and second-order asymptotics of classical communication. First, as already mentioned, the achievability proof relies on PPM codes [46] instead of iid random codes. Second, unlike the analysis of the second-order asymptotics of reliable and secure communications [35, 47], we have to deal here with parameters capturing reliability and covertness, with the latter appearing in the first-order asymptotics; more specifically, the optimal coding scheme identified in [13] exploits a code with a bin structure, in which each bin forms a reliability code for the legitimate channel indexed by the secret key while the overall code forms a resolvability code for the adversary's channel. To prove the existence of a code with the desired characteristics, we expurgate a random code after resorting to concentration of measure inequalities, such as McDiarmid's inequality, and carefully analyzing the probability of error. We point out that our current results

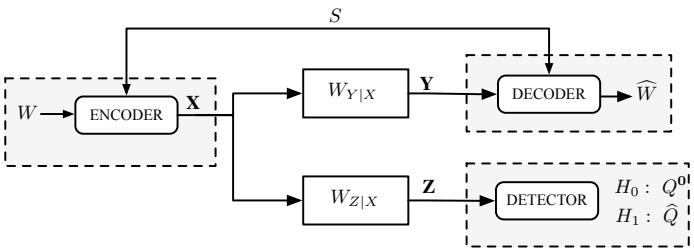

Figure 2.3.1: Model of covert communications over noisy channels

only identify the second-order asymptotics for the number of transmitted message bits and do not characterize the second-order asymptotics for the number of key bits.

2.3 Problem Formulation

We consider the point-to-point covert communication setup described in Section 1.1.1 and depicted in Fig. 2.3.1. Alice (the legitimate transmitter) is connected to Bob (the legitimate receiver) and Willie (the warden) through the channels $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ and $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$, respectively. We recall that the transmission of $0 \in \mathcal{X}$ indicates no communication and we denote $P_x \triangleq W_{Y|X=x}$ and $Q_x \triangleq W_{Z|X=x}$ for all $x \in \mathcal{X}$, as well as $P^{\mathbf{x}} \triangleq P_{x_1} \otimes \cdots \otimes P_{x_n}$ and $Q^{\mathbf{x}} \triangleq Q_{x_1} \otimes \cdots \otimes Q_{x_n}$ for all $\mathbf{x} \in \mathcal{X}^n$. Alice wishes to transmit a message uniformly distributed in a set $\mathcal{M}$ to Bob with the help of a secret shared key uniformly distributed over $\mathcal{K}$. Let $\mathbf{Z}$ denote Willie's observations in n channel uses. The distribution of $\mathbf{Z}$ is $Q^{\mathbf{0}}$ where $\mathbf{0}$ denotes the all-zero sequence in $\mathcal{X}^n$. We also denote by $\widehat{Q}$ the distribution of $\mathbf{Z}$ when there exists communication. Motivated by the discussion in Section 1.1.1, we measure the covertness as a "distance," denoted by $d(\cdot,\cdot)$, between the distribution of $\mathbf{Z}$ with and without communication, i.e., $d(\widehat{Q}, Q^{\mathbf{0}})$. Here is the formal definition of the covert communication code.

Definition 2. *A code $\mathcal{C}$ with message set $\mathcal{M}$, key set $\mathcal{K}$, and block-length n consists of an encoder $f : \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n$ and a decoder $\phi : \mathcal{K} \times \mathcal{Y}^n \to \mathcal{M}$. It is called an $(M, K, \epsilon, \delta)^d_{\text{avg}}$*

code if

$$|\mathcal{M}| \geqslant M, \quad |\mathcal{K}| \leqslant K, \quad d(\widehat{Q}, Q^{\mathbf{0}}) \leqslant \delta, \tag{2.1}$$

$$\frac{1}{MK} \sum_{m \in \mathcal{M}, k \in \mathcal{K}} P^{f(m,k)}(\{\mathbf{y} : \phi(k, \mathbf{y}) = m\}) \geqslant 1 - \epsilon. \tag{2.2}$$

Let $M^*(\epsilon, \delta)^d_{\text{avg}} \triangleq \sup\big\{M : \exists (M, K, \epsilon, \delta)^d_{\text{avg}}$ *code*$\big\}$.

We similarly define an $(M, K, \epsilon, \delta)^d_{\max}$ *code and* $M^*(\epsilon, \delta)^d_{\max}$ *by replacing* (2.2) *by*

$$\max_{m \in \mathcal{M}, k \in \mathcal{K}} P^{f(m,k)}(\{\mathbf{y} : \phi(k, \mathbf{y}) = m\}) \geqslant 1 - \epsilon. \tag{2.3}$$

In particular, we study three distances, namely, the relative entropy $d(P, Q) = \mathbb{D}(P\|Q)$*, vartional distance* $d(P, Q) = \frac{1}{2}\|P - Q\|_1$*, and the minimum probability of missed detection at a fixed false alarm probability* $d(P, Q) = 1 - \alpha - \beta_\alpha(Q, P)$.

Remark 3. *Most of the previous studies of covert communication [12, 13] considered the* average *probability of error; we study here both the average and the maximum probability of error. The latter constraint captures a pragmatic requirement and is critical to our second-order converse argument. As shown in Appendix A, analyzing the average probability of error is a different and (we believe) substantially more difficult endeavor for the non-vanishing probability of error since there is no strong converse for the reliability parameter. This intriguing behavior appears because one can slightly increase the average probability of error without any effect on covertness by simply adding all-zero codewords to the codebook. This, of course, leads to what one would consider as "bad codes," which we avoid by focusing on the maximal probability of error.*

Remark 4. *The three covertness metrics discussed earlier have slightly different operational meanings. When enforcing that the probability of missed detection satisfy* $\beta_\alpha(Q^{\mathbf{0}}, \widehat{Q}) \geqslant 1 - \alpha - \delta$ *for a fixed probability of false alarm* α*, one implicitly assumes that the adversary optimizes its detector at a specific point of the ROC curve with known probability*

of false alarm. In contrast, when enforcing $\frac{1}{2}\|\widehat{Q} - Q^{\mathbf{0}}\|_1 \leqslant \delta$*, since* $\alpha + \beta_\alpha(Q^{\mathbf{0}}, \widehat{Q}) \geqslant 1 - \frac{1}{2}\|\widehat{Q} - Q^{\mathbf{0}}\|_1$ *for* any *probability of false alarm* α*, one essentially wishes to enforce covertness irrespective of the exact operating point on the adversary's ROC curve. Finally, since* $\frac{1}{2}\|\widehat{Q} - Q^{\mathbf{0}}\|_1^2 \leqslant \frac{1}{2}\mathbb{D}(\widehat{Q}\|Q^{\mathbf{0}})$ *by Pinsker's inequality, the constraint* $\mathbb{D}(\widehat{Q}\|Q^{\mathbf{0}}) \leqslant \delta$ *is more stringent than when using variational distance, but only provides a loose proxy for constraining the ROC curve since Pinsker's inequality is not tight.*[1] *The relations between the covertness metrics immediately lead to the following ordering of the maximum number of covert bits.*

$$M^*(\epsilon, \delta)^D_{\text{avg}} \leqslant M^*(\epsilon, \sqrt{\frac{\delta}{2}})^V_{\text{avg}} \leqslant \min_{\alpha \in]0,1[} M^*(\epsilon, \sqrt{\frac{\delta}{2}})^{\beta_\alpha}_{\text{avg}} \tag{2.4}$$

$$M^*(\epsilon, \delta)^D_{\max} \leqslant M^*(\epsilon, \sqrt{\frac{\delta}{2}})^V_{\max} \leqslant \min_{\alpha \in]0,1[} M^*(\epsilon, \sqrt{\frac{\delta}{2}})^{\beta_\alpha}_{\max}, \tag{2.5}$$

where D *denotes the relative entropy,* V *denotes varitional distance and* β_α *denotes the distance* $1 - \beta_\alpha(P, Q) - \alpha$*.*

2.4 Main Results

To simplify our results, we restrict the channels to binary-input channels, i.e., $\mathcal{X} = \{0, 1\}$. We also assume that $Q_1 \ll Q_0$, $P_1 \ll P_0$ and $Q_1 \neq Q_0$. These assumptions are necessary for our results to hold. In fact, without $Q_1 \ll Q_0$, covert communication is impossible as some input symbols detect the use of the non-innocent symbol with probability one. At the other extreme, if $Q_1 = Q_0$ then no detector can distinguish the use of innocent and non-innocent symbols so that the problem reduces to classical communication. Finally, without $P_1 \ll P_0$, Alice and Bob have an unfair advantage that allows them to exchange $\omega(\sqrt{n})$ bits covertly [13, Theorem 7]. Our main results in Theorem 7, Theorem 8, and Theorem 9 characterize the maximum number of reliable and covert bits under the above assumptions. In

[1] See also the discussion in [13, Appendix A].

all cases, the first and second terms behave as $\Theta\left(n^{\frac{1}{2}}\right)$ and $\Theta\left(n^{\frac{1}{4}}\right)$, respectively, as expected; nevertheless, the constant behind $\Theta(\cdot)$ is metric-specific.

Theorem 7. *Let $d(P,Q) = \mathbb{D}(P\|Q)$, $\epsilon \in]0,1[$ and $\delta > 0$ be fixed, and $\omega \triangleq \sqrt{\frac{2\delta}{\chi_2(Q_1\|Q_0)}}$. We then have*

$$\log M^*(\epsilon,\delta)^d_{\max} = \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n). \tag{2.6}$$

This optimal number of message bits is obtained with a first-order optimal number of key bits

$$\log K = \max(0, \omega(\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0))n^{\frac{1}{2}} + O(n^{\frac{1}{4}}\sqrt{\log n})), \tag{2.7}$$

In addition,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{\log M^*(\epsilon,\delta)^d_{\text{avg}}}{\sqrt{n}} = \frac{\log M^*(\epsilon,\delta)^d_{\max}}{\sqrt{n}} = \omega\mathbb{D}(P_1\|P_0). \tag{2.8}$$

Theorem 8. *Let $d(P,Q) = \frac{1}{2}\|P-Q\|_1$, $\epsilon \in]0,1[$ and $\delta \in]0,1[$ be fixed, $\Gamma \triangleq \mathbb{Q}^{-1}\left(\frac{1-\delta}{2}\right)$, and $\omega \triangleq \frac{2\Gamma}{\sqrt{\chi_2(Q_1\|Q_0)}}$. We have*

$$\log M^*(\epsilon,\delta)^d_{\max} \leqslant \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n), \tag{2.9}$$

and

$$\begin{aligned}\log M^*(\epsilon,\delta)^d_{\max} \geqslant\ & \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} \\ & - \left(\sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon) + \frac{2\sqrt{\pi}e^{\frac{\Gamma^2}{2}}\mathbb{D}(P_1\|P_0)}{\sqrt{\Gamma}\chi_2(Q_1\|Q_0)^{\frac{1}{4}}}\right)n^{\frac{1}{4}} + O(\log n).\end{aligned} \tag{2.10}$$

This number of message bits is obtained with a number of key bits

$$\log K = \max(0, \omega(\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0))n^{\frac{1}{2}} + O(n^{\frac{1}{4}}\sqrt{\log n})). \tag{2.11}$$

In addition,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{\log M^*(\epsilon,\delta)^d_{\text{avg}}}{\sqrt{n}} = \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{\log M^*(\epsilon,\delta)^d_{\max}}{\sqrt{n}} = \omega\mathbb{D}(P_1\|P_0). \tag{2.12}$$

Theorem 9. *Let* $d(P,Q) = 1-\alpha-\beta_\alpha(Q,P)$ *and* $\epsilon \in]0,1[$, $\alpha \in]0,1[$, *and* $\delta \in]0, 1-\alpha[$ *be fixed. Let* $\Lambda \triangleq \mathbb{Q}^{-1}(1-\alpha-\delta)$, $\Upsilon \triangleq \mathbb{Q}^{-1}(\alpha)$, *and* $\omega \triangleq \frac{\Lambda+\Upsilon}{\sqrt{\chi_2(Q_1\|Q_0)}}$. *We have*

$$\log M^*(\epsilon,\delta)^d_{\max} \leqslant \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n), \tag{2.13}$$

and

$$\log M^*(\epsilon,\delta)^d_{\max} \geqslant \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \left(\sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon) + \frac{\sqrt{2\pi}\left(e^{\frac{\Gamma^2}{2}} + e^{\frac{\Upsilon^2}{2}}\right)\mathbb{D}(P_1\|P_0)}{\sqrt{\Lambda+\Upsilon}\chi_2(Q_1\|Q_0)^{\frac{1}{4}}}\right)n^{\frac{1}{4}} + O(\log n). \tag{2.14}$$

This number of message bits is obtained with a number of key bits

$$\log K = \max(0, \omega(\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0))n^{\frac{1}{2}} + O(n^{\frac{1}{4}}\sqrt{\log n})). \tag{2.15}$$

In addition,

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{\log M^*(\epsilon,\delta)^d_{\text{avg}}}{\sqrt{n}} = \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{\log M^*(\epsilon,\delta)^d_{\max}}{\sqrt{n}} = \omega\mathbb{D}(P_1\|P_0). \tag{2.16}$$

We point out that the second order asymptotics in Theorem 8 and Theorem 9 are loose because the parameters Γ, Λ, and Υ may be very small; in particular, because of the in-

equalities (2.4), Theorem 7 is sometimes a tighter lower bound for small values of n. We conjecture that the upper-bounds in (2.9) and (2.13) can be achieved, although we could not establish it with our current proof techniques.

We illustrate the results of the three theorems with a simple numerical example. We consider the situation in which $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ and $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ are BSCs with cross-over probability $p_m = 0.11$ and $p_w = 0.45$, respectively, and $\epsilon = 10^{-3}$, $\delta = 10^{-2}$, and $\alpha = 0.2$. Since the convergence to the asymptotic limit is slow (on the order of $\Theta(n^{-\frac{1}{4}})$), we use a log scale for the blocklength. As shown in Fig. 2.4.1, the choice of the covertness metric results in different number of bits, which of course raises the question of which number to settle on.

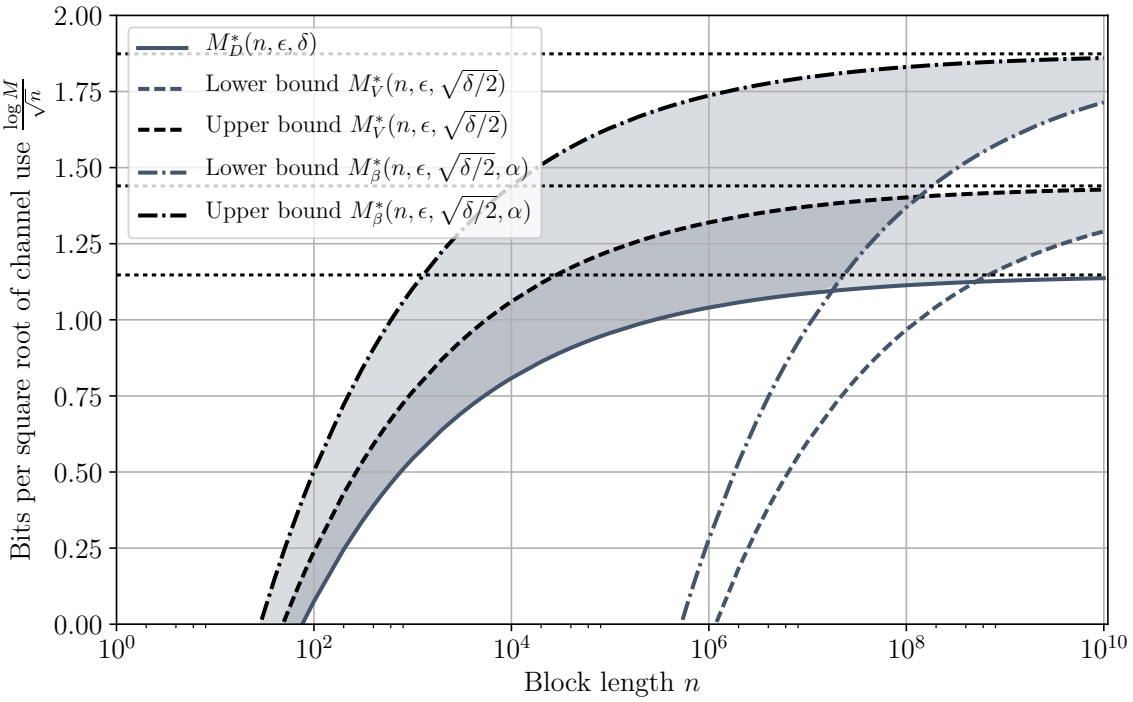

Figure 2.4.1: Second-order approximation of maximum number of covert and reliable bits as a function of blocklength. Both DMCs are BSCs with cross-over probability $p_m = 0.11$ and $p_w = 0.45$, respectively, and $\epsilon = 10^{-3}$, $\delta = 10^{-2}$, and $\alpha = 0.2$. The dotted horizontal lines indicate the optimal first-order asymptotics.

2.5 Results for a General Distance

2.5.1 Information Specturm Divergence

Our achievability and converse results will be stated in terms of *information spectrum divergence* [48] defined as follows.

Definition 3. *Let μ and ν be two probability measures over the same space such that $\nu \ll \mu$. We define*

$$\mathbb{D}_s^{\epsilon}\left(\mu\|\nu\right) \triangleq \sup\left\{\gamma : \mathbb{P}_{\mu}\left(\log\frac{d\mu}{d\nu} \leqslant \gamma\right) \leqslant \epsilon\right\}, \tag{2.17}$$

where $\frac{d\mu}{d\nu}$ denote the Radon-Nykonym derivative.

We now review some useful properties.

1. By the continuity of probability measures, we have

$$\mathbb{P}_{\mu}\left(\log\frac{d\mu}{d\nu} < \mathbb{D}_s^{\epsilon}\left(\mu\|\nu\right)\right) \leqslant \epsilon. \tag{2.18}$$

2. For any third probability measure ω, we have

$$\mathbb{D}_s^{\epsilon}\left(\mu\otimes\omega\|\nu\otimes\omega\right) = \mathbb{D}_s^{\epsilon}\left(\mu\|\nu\right). \tag{2.19}$$

3. Given that $\mathbb{E}_{\mu}\left(\left|\log\frac{d\mu}{d\nu}\right|^i\right)$ is finite for $i = 1, 2, 3$, we have by Theorem 6 in Chapter 1,

$$\mathbb{D}_s^{\epsilon}\left(\mu^{\otimes n}\|\nu^{\otimes n}\right) \geqslant n\mathbb{D}(\mu\|\nu) - \sqrt{nV(\mu\|\nu)}\mathbb{Q}^{-1}\left(\epsilon - \frac{T(\mu\|\nu)}{\sqrt{n}V^{\frac{3}{2}}(\mu\|\nu)}\right), \tag{2.20}$$

and

$$\mathbb{D}_s^{\epsilon}\left(\mu^{\otimes n}\|\nu^{\otimes n}\right) \leqslant n\mathbb{D}(\mu\|\nu) - \sqrt{nV(\mu\|\nu)}\mathbb{Q}^{-1}\left(\epsilon + \frac{T(\mu\|\nu)}{\sqrt{n}V^{\frac{3}{2}}(\mu\|\nu)}\right). \tag{2.21}$$

4. By Theorem 3 in Chapter 1, we have

$$\mathbb{D}_s^{1-\epsilon}\left(\mu^{\otimes n}\|\nu^{\otimes n}\right) \geqslant n\mathbb{D}(\mu\|\nu) + \sqrt{\frac{n\log\frac{1}{\epsilon}}{2}}\left\|\log\frac{d\mu}{d\nu}\right\|_{\mu,\infty}. \tag{2.22}$$

5. We now restrict ourselves to discrete probabilities for simplicity. For any PMF P_X and for all channels $W_{Y|X}$ and $V_{Y|X}$, we have [48]

$$\mathbb{D}_s^{\epsilon}\left(P_X \otimes W_{Y|X}\|P_X \otimes V_{Y|X}\right) \leqslant \sup_{x\in\text{support}(P_X)} \mathbb{D}_s^{\epsilon}\left(W_{Y|X=x}\|V_{Y|X=x}\right). \tag{2.23}$$

2.5.2 Achievability

Finite-Length Achievability Results

To establish achievability results, we need a "continuity" condition on the metric d for the following reason. For a random codebook whose codewords are generated independently from a distribution $P_{\mathbf{X}}$, one can expect that if the size of the codebook is large enough the induced distribution by the codebook is closed to $W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}$. We can also choose $P_{\mathbf{X}}$ such that $d(W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}, Q^{\mathbf{0}})$ is small. The continuity then allows us to bound the distance between the induced distribution by the codebook and $Q^{\mathbf{0}}$.

Definition 4. *Let $L : \mathbb{R} \to \mathbb{R}$ be a function that maps positive numbers to positive numbers. We say that a distance d is L-continuous at a distribution Q if*

$$\forall P_1, P_2 : |d(P_1, Q) - d(P_2, Q)| \leqslant L(\|P_1 - P_2\|_1) \tag{2.24}$$

The following theorem provides an achievability result for maximum probability of error.

Theorem 10. *Let d be L-continuous at $Q^{\mathbf{0}}$. Let $P_{\mathbf{X}}$ be a probability distribution over $\mathcal{X}^n$.*

Let $\mathcal{M}$ and $\mathcal{K}$ be two sets and ϵ, δ', ϵ', δ'', η, p, λ be real numbers such that

$$\log|\mathcal{M}| \leqslant \inf_{x\in\mathrm{support}(P_{\mathbf{X}})} \mathbb{D}_s^{\epsilon-\epsilon'}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{(1-\eta)(1-p)\epsilon'}{\chi_2\left(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}}\right)+1} \tag{2.25}$$

$$\log|\mathcal{M}| + \log|\mathcal{K}| \geqslant \mathbb{D}_s^{1-\delta'+\delta''}\left(P_{\mathbf{X}}\otimes W_{Z|X}^{\otimes n}\|P_{\mathbf{X}}\otimes Q^{\mathbf{0}}\right) + \log\frac{1}{4\delta''^2} \tag{2.26}$$

$$p > \exp\left(-2|\mathcal{M}|\lambda^2\right) \tag{2.27}$$

There exists an $(\eta|\mathcal{M}|, |\mathcal{K}|, \epsilon, L(\delta'+\lambda+1-\eta)+d(W_{Z|X}^{\otimes n}\circ P_{\mathbf{X}}, Q^{\mathbf{0}}))_{\max}^d$ code.

Proof. See Appendix 2.B. □

We provide here a high-level sketch of the proof. Following [13], the coding scheme in the achievability proof consists of $|\mathcal{M}|\times|\mathcal{K}|$ randomly generated codewords $\mathbf{x}_{sw}$ with $s\in\mathcal{K}$ and $w\in\mathcal{M}$. The code is designed such that the following two properties hold:

(P1) the entire codebook $\{\mathbf{x}_{sw}\}_{(s,w)\in\mathcal{K}\times\mathcal{M}}$ forms a resolvability code for the channel $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ approximating $W_{Z|X}^{\otimes n}\circ P_{\mathbf{X}}$;

(P2) for every $s\in\mathcal{K}$, with probability at least p (with respect to the random coding), there exists a subset $\widetilde{\mathcal{M}}\subset\mathcal{M}$ with $\left|\widetilde{\mathcal{M}}\right| \geqslant \eta|\mathcal{M}|$ and sub-codebook $\{\mathbf{x}_{sw}\}_{w\in\widetilde{\mathcal{M}}}$ forms a reliability code for the channel $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ with a *maximum* probability of error ϵ.

One cannot ensure (P1) and (P2) by a standard expurgation argument, since one should expurgate a small number of codewords to ensure that the induced output probability of the expurgated codebook does not change much. We address this issue by a careful analysis of the probability of error of a random codebook.

Remark 5. *The extra term $\chi_2\left(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}}\right)+1$ in* (2.25)*, which does not appear in [35, Theorem 18], is the penalty of using $P^{\mathbf{0}}$ instead of the "true" probability distribution of $\mathbf{Y}$ in our decoding rule. This sub-optimal bound yields the same first- and second-*

order asymptotics for covert communications, although it could certainly be improved for a finite-length analysis; this choice, however, simplifies our calculations.

PPM Distribution

We now study the specific "PPM distribution," which was used in [46] for covert communication; we subsequently use this distribution to generate random codes and combine them with Theorem 10.

Definition 5 ((n,ℓ)-PPM covert distribution)**.** *Given $\mathcal{X}=\{0,1\}$ and $n \geqslant \ell \geqslant 1$, we define the distribution $P^{n,\ell}_{\mathbf{X},\text{PPM}}$ on $\mathcal{X}^n$ as follows. If $n = m\ell + r$ for $0 \leqslant r < \ell$, we define*

$$\mathcal{B} \triangleq \left\{ \mathbf{x} \in \mathcal{X}^n : \forall i \in [\![1,\ell]\!] : \sum_{j=(i-1)m+1}^{im} x_j = 1 \text{ and } \sum_{j=\ell m+1}^{n} x_j = 0 \right\}. \tag{2.28}$$

We then set $P^{n,\ell}_{\mathbf{X},\text{PPM}}$ as the uniform distribution on $\mathcal{B}$. We further define $P^{n,\ell}_{\mathbf{Y},\text{PPM}} \triangleq W^{\otimes n}_{Y|X} \circ P^{n,\ell}_{\mathbf{X},\text{PPM}}$ and $P^{n,\ell}_{\mathbf{Z},\text{PPM}} \triangleq W^{\otimes n}_{Z|X} \circ P^{n,\ell}_{\mathbf{X},\text{PPM}}$.

The PPM distribution $P^{n,\ell}_{\mathbf{X},\text{PPM}}$ may be viewed as a way to generate structured constant composition codewords; specifically, every codeword generated according to $P^{n,\ell}_{\mathbf{X},\text{PPM}}$ contains exactly ℓ 1-symbols, each located in a window of size m, the quotient of the Euclidean division of n by ℓ. The structure of the PPM covert distribution plays a pivotal role in our analysis as 1) all sequences $\mathbf{x} \in \text{support}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}}\right)$ have the same weight (unlike iid distribution), which is essential to bound $\inf_{x\in\text{support}(P_{\mathbf{X}})} \mathbb{D}_s^{\epsilon-\epsilon'}\left(P^{\mathbf{x}} \| Q_{\mathbf{Y}}\right)$ in (2.25); 2) the structure $P^{n,\ell}_{\mathbf{X},\text{PPM}} = \left(P^{m,1}_{\mathbf{X},\text{PPM}}\right)^{\otimes \ell} \otimes P_0^{\otimes r}$ simplifies the calculation of $d(P^{n,\ell}_{\mathbf{Z},\text{PPM}}, Q^{\mathbf{0}})$.

We devote the remainder of this section to the development of properties of PPM distribution, which are geared towards the use of Theorem 10.

Lemma 8. *We have for* $\epsilon - \epsilon' - \frac{6T(P_1\|P_0)}{\sqrt{\ell}V^{\frac{3}{2}}(P_1\|P_0)} > 0$*,*

$$\inf_{x\in\text{support}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}}\right)} \mathbb{D}_s^{\epsilon-\epsilon'}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right)$$
$$\geqslant \ell\mathbb{D}(P_1\|P_0) - \sqrt{\ell V(P_1\|P_0)}\mathbb{Q}^{-1}\left(\epsilon - \epsilon' - \frac{6T(P_1\|P_0)}{\sqrt{\ell}V^{\frac{3}{2}}(P_1\|P_0)}\right) \quad (2.29)$$

We further have

$$\chi_2\left(P^{n,\ell}_{\mathbf{Y},\text{PPM}}\|P^{\mathbf{0}}\right) + 1 \leqslant \exp\left(\frac{\ell(\ell+1)}{n}\chi_2(P_1\|P_0)\right) \quad (2.30)$$

$$\mathbb{D}_s^{1-\delta}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}}\otimes W_{Z|X}^{\otimes n}\|P^{n,\ell}_{\mathbf{X},\text{PPM}}\otimes Q^{\mathbf{0}}\right) \leqslant \ell\mathbb{D}(Q_1\|Q_0) + \sqrt{\frac{\ell\log\frac{1}{\delta}}{2}}\sup_{z:Q_1(0)>0}\left|\frac{Q_1(z)}{Q_0(z)}\right| \quad (2.31)$$

Proof. Note first that for any $\mathbf{x}$ with $\text{wt}(\mathbf{x}) = \ell$, we have

$$\mathbb{D}_s^{\epsilon-\epsilon'}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) \overset{(a)}{=} \mathbb{D}_s^{\epsilon-\epsilon'}\left(P_1^{\otimes\ell}\|P_0^{\otimes\ell}\right) \quad (2.32)$$
$$\overset{(b)}{\geqslant} \ell\mathbb{D}(P_1\|P_0) - \sqrt{\ell V(P_1\|P_0)}\mathbb{Q}^{-1}\left(\epsilon - \epsilon' - \frac{6T(P_1\|P_0)}{\sqrt{\ell}V^{\frac{3}{2}}(P_1\|P_0)}\right), \quad (2.33)$$

where (a) follows from (2.19) and (b) follows from (2.20). Since we have $\text{wt}(\mathbf{x}) = \ell$ for all $\mathbf{x} \in \text{support}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}}\right)$, we have (2.29).

To prove (2.30), we define $m \triangleq \lfloor n/\ell \rfloor$. We then have

$$\chi_2(P^{n,\ell}_{\mathbf{Y},\text{PPM}} \| P^{\mathbf{0}}) + 1 = \chi_2(P^{m\ell,\ell}_{\mathbf{Y},\text{PPM}} \| P^{\otimes m\ell}) + 1 \tag{2.34}$$

$$= \left(\chi_2\left(P^{m,1}_{\mathbf{Y},\text{PPM}} \| P^{\otimes m}\right) + 1\right)^{\ell} \tag{2.35}$$

$$= \left(\sum_{\mathbf{y}} P_0^{\otimes m}(\mathbf{y}) \left(\frac{P^{m,1}_{\mathbf{Y},\text{PPM}}(\mathbf{y})}{P_0^{\otimes m}(\mathbf{y})}\right)^2\right)^{\ell} \tag{2.36}$$

$$= \left(\sum_{\mathbf{y}} P_0^{\otimes m}(\mathbf{y}) \left(\frac{1}{m}\sum_{i=1}^{m} \frac{P_1(y_i)}{P_0(y_i)}\right)^2\right)^{\ell} \tag{2.37}$$

$$= \left(\frac{1}{m^2}\left(m(m-1) + m(\chi_2(P_1\|P_0) + 1)\right)\right)^{\ell} \tag{2.38}$$

$$= \left(1 + \frac{\chi_2(P_1\|P_0)}{m}\right)^{\ell} \tag{2.39}$$

$$\leqslant \exp\left(\frac{\ell(\ell+1)}{n}\chi_2(P_1\|P_0)\right). \tag{2.40}$$

Finally, note that

$$\mathbb{D}_s^{1-\delta}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}} \otimes W_{Z|X}^{\otimes n} \| P^{n,\ell}_{\mathbf{X},\text{PPM}} \otimes Q^{\mathbf{0}}\right) \overset{(a)}{\leqslant} \sup_{x \in \text{support}\left(P^{n,\ell}_{\mathbf{X},\text{PPM}}\right)} \mathbb{D}_s^{1-\delta}\left(Q^{\mathbf{x}} \| Q^{\mathbf{0}}\right) \tag{2.41}$$

$$\overset{(b)}{=} \mathbb{D}_s^{1-\delta}\left(Q_1^{\otimes \ell} \| Q_0^{\otimes \ell}\right) \tag{2.42}$$

$$\overset{(c)}{\leqslant} \ell \mathbb{D}(Q_1\|Q_0) + \sqrt{\frac{\ell \log \frac{1}{\delta}}{2}} \sup_{z: Q_1(0)>0} \left|\frac{Q_1(z)}{Q_0(z)}\right|, \tag{2.43}$$

where (a) follows from (2.23), (b) follows from (2.19), and (c) follows from (2.22). □

We now bound $d(P^{n,\ell}_{\mathbf{Z},\text{PPM}}, Q^{\mathbf{0}})$ for the three metrics.

Lemma 9. *Let n and ℓ be positive integers with $m \triangleq \lfloor n/\ell \rfloor$ large enough and $\ell = \Theta(m)$. We then have*

$$\mathbb{D}\left(P^{n,\ell}_{\mathbf{Z},\text{PPM}} \| Q^{\mathbf{0}}\right) \leqslant \frac{\ell^2}{2n}\chi_2(Q_1\|Q_0) + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.44}$$

Furthermore, we have

$$\frac{1}{2}\left\|P^{n,\ell}_{\mathbf{Z},\text{PPM}} - Q^{\mathbf{0}}\right\|_1 \leqslant 1 - 2\mathbb{Q}\left(\frac{\ell}{2}\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}}\right) + \frac{2}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right), \tag{2.45}$$

$$\beta_\alpha\left(Q^{\mathbf{0}}, P^{n,\ell}_{\mathbf{Z},\text{PPM}}\right) \geqslant \mathbb{Q}\left(\ell\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}} - \mathbb{Q}^{-1}\left(\alpha + \frac{1}{\sqrt{\ell}}\right)\right) - \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.46}$$

Proof. See Appendix 2.C. □

2.5.3 Converse

We now develop a generic converse for a distance d. The following result states that the number of covert and reliable bits that one can transmit may be characterized by establishing an upper bound on the weight of codewords. Such upper bounds are metric-specific, and we develop them in Section 2.6.

Definition 6. *For any distance d and real numbers $w > 0$ and $\delta > 0$, let*

$$\lambda_d(w,\delta) \triangleq \inf_{P_{\mathbf{X}}: d(W^{\otimes n}_{Z|X}\circ P_{\mathbf{X}}, Q^{\mathbf{0}}) \leqslant \delta} \mathbb{P}_{P_{\mathbf{X}}}(wt(\mathbf{X}) \leqslant w). \tag{2.47}$$

Theorem 11. *Suppose that there exists constants ω and C depending only on Q_1, Q_0 and δ such that for all n large enough and γ small enough $\lambda_d((\omega - C\gamma)n^{\frac{1}{2}}, \delta) \geqslant \gamma$. We then have*

$$\log M^*(\epsilon,\delta)^d_{\max} \leqslant \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n), \tag{2.48}$$

and

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{M^*(\epsilon,\delta)^d_{\text{avg}}}{\sqrt{n}} \leqslant \omega\mathbb{D}(P_1\|P_0). \tag{2.49}$$

Proof. See Appendix 2.D. □

2.6 Covert Communication with Specific Covertness Metrics

We now leverage the general results established in Section 2.5 and specialize them to study three covertness metrics: relative entropy (Subsection 2.6.1), variational distance (Subsection 2.6.2), and probability of missed detection (Subsection 2.6.3). As alluded to earlier, all that needs to be done is: (i) establish that the metric under consideration is a quasi metric, as defined at the beginning of Section 2.5; (ii) verify that the conditions of Theorem 10 and Theorem 12 are satisfied.

2.6.1 Covertness in Relative Entropy

In this subsection, we prove Theorem 7.

Lemma 10. *Suppose $\mathcal{Z}$ is finite and let $P_{\mathbf{Z}}$ and $Q_{\mathbf{Z}}$ be two PMFs over $\mathcal{Z}^n$ such that $\frac{1}{2}\|P_{\mathbf{Z}} - Q_{\mathbf{Z}}\| \leqslant \epsilon$. We then have*

$$\left|\mathbb{D}\big(P_{\mathbf{Z}}\|Q^{\mathbf{0}}\big) - \mathbb{D}\big(Q_{\mathbf{Z}}\|Q^{\mathbf{0}}\big)\right| \leqslant \epsilon \log\left(\frac{|\mathcal{Z}|}{\left(\max_{z\in\mathcal{Z}} Q_0(z)\right)^2}\right) n + \mathbb{H}_b\left(\epsilon\right). \tag{2.50}$$

Proof. Note that

$$\left|\mathbb{D}\big(P_{\mathbf{Z}}\|Q^{\mathbf{0}}\big) - \mathbb{D}\big(Q_{\mathbf{Z}}\|Q^{\mathbf{0}}\big)\right| = \left(H(P_{\mathbf{Z}}) - H(Q_{\mathbf{Z}}) + \sum_{\mathbf{z}} \left(P_{\mathbf{Z}}(\mathbf{z}) - Q_{\mathbf{Z}}(\mathbf{z})\right) \log \frac{1}{Q^{\mathbf{0}}(\mathbf{z})}\right) \tag{2.51}$$

$$\leqslant \left|H(P_{\mathbf{Z}}) - H(Q_{\mathbf{Z}})\right| + \sum_{\mathbf{z}} \left|P_{\mathbf{Z}}(\mathbf{z}) - Q_{\mathbf{Z}}(\mathbf{z})\right| \log \frac{1}{Q^{\mathbf{0}}(\mathbf{z})} \tag{2.52}$$

$$\overset{(a)}{\leqslant} \epsilon \log(|\mathcal{Z}|)n + \mathbb{H}_b\left(\epsilon\right) + \sum_{\mathbf{z}} \left|P_{\mathbf{Z}}(\mathbf{z}) - Q_{\mathbf{Z}}(\mathbf{z})\right| \log \frac{1}{Q^{\mathbf{0}}(\mathbf{z})} \tag{2.53}$$

$$\leqslant \epsilon \log(|\mathcal{Z}|)n + \mathbb{H}_b\left(\epsilon\right) + 2\epsilon \log \frac{1}{\max_z Q_0(z)} n, \tag{2.54}$$

Parameter	Value
ℓ_n	$\lfloor \omega\sqrt{n} - t \rfloor$
$P_{\mathbf{X}}^n$	$P_{\mathbf{X},\text{PPM}}^{n,\ell_n}$
ϵ_n'	$\frac{1}{\sqrt{\ell_n}}$
η_n	$1 - \frac{1}{n^2}$
p_n	$\frac{1}{2}$
δ_n'	$\frac{1}{n^2}$
δ_n''	$\frac{1}{2n^2}$
λ_n	$\frac{1}{n^2}$

Figure 2.6.1: The parameters in the proof of Theorem 7

where (a) follows from Fannes's inequality. □

Achievability of Theorem 7. Let ϵ, δ, and ω be as in Theorem 7. To use Theorem 10, we choose parameters $\ell_n, P_{\mathbf{X}}^n, \epsilon_n', \eta_n, p_n, \delta_n', \delta_n'', \lambda_n$ as in Table 2.6.1 where t will be determined later. We then choose $\mathcal{M}_n$ and $\mathcal{K}_n$ such that

$$\log|\mathcal{M}_n| = \inf_{\mathbf{x}\in\text{support}\left(P_{\mathbf{X}}^n\right)} \mathbb{D}_s^{\epsilon-\epsilon_n'}\left(P^{\mathbf{x}}\|Q_{\mathbf{Y}}\right) + \log\frac{(1-\eta_n)(1-p_n)\epsilon_n'}{\chi_2(P_{\mathbf{Y},\text{PPM}}^{n,\ell_n}\|P^{\mathbf{0}})+1} \tag{2.55}$$

$$\overset{(a)}{\geqslant} \ell_n\mathbb{D}(P_1\|P_0) - \sqrt{\ell_n V(P_1\|P_0)}\mathbb{Q}^{-1}\left(\epsilon-\epsilon_n'-\frac{6T(P_1\|P_0)}{\sqrt{\ell_n}V^{\frac{3}{2}}(P_1\|P_0)}\right) + \log\frac{(1-\eta_n)(1-p_n)\epsilon_n'}{\chi_2(P_{\mathbf{Y},\text{PPM}}^{n,\ell_n}\|P^{\mathbf{0}})+1} \tag{2.56}$$

$$\overset{(b)}{\geqslant} \ell_n\mathbb{D}(P_1\|P_0) - \sqrt{\ell_n V(P_1\|P_0)}\mathbb{Q}^{-1}\left(\epsilon-\epsilon_n'-\frac{6T(P_1\|P_0)}{\sqrt{\ell_n}V^{\frac{3}{2}}(P_1\|P_0)}\right) + \log\frac{(1-\eta_n)(1-p_n)\epsilon_n'}{\exp\left(\ell_n(\ell_n+1)\chi_2(P_1\|P_0)\right))} \tag{2.57}$$

$$= \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n) + O(1) \tag{2.58}$$

and

$$\log|\mathcal{M}_n| + \log|\mathcal{K}_n| \triangleq \max(\log|\mathcal{M}_n|, \mathbb{D}_s^{1-\delta_n'+\delta_n''}\left(P_{\mathbf{X}}^n \otimes W_{Z|X}^{\otimes n} \| P_{\mathbf{X}}^n \otimes Q_{\mathbf{Z}}\right) + \log\frac{1}{4\delta_n''^2}) \tag{2.59}$$

$$\overset{(a)}{\leqslant} \max(\log|\mathcal{M}_n|, \ell_n \mathbb{D}(Q_1\|Q_0) + \sqrt{\frac{\ell_n \log\frac{1}{\delta_n'}}{2}} \sup_{z:Q_1(z)>0}\left|\log\frac{Q_1(z)}{Q_0(z)}\right| + \log\frac{1}{4\delta_n''^2}) \tag{2.60}$$

$$= \max(\log|\mathcal{M}_n|, \omega_n \mathbb{D}(Q_1\|Q_0) n^{\frac{1}{2}} + O(n^{\frac{1}{4}}\sqrt{\log n}). \tag{2.61}$$

We also have $p_n \geqslant \exp\left(-2|\mathcal{M}_n|\,\lambda_n^2\right)$ as p_n is constant and $\exp\left(-2|\mathcal{M}_n|\,\lambda_n^2\right)$ tends to zero. We additionally choose t such that $\mathbb{D}\Big(P_{\mathbf{Z},\text{PPM}}^{n,\ell_n}\|Q^{\mathbf{0}}\Big) \leqslant \delta - \frac{1}{\sqrt{n}}$. This is possible because of the definition of ℓ_n and (2.44). Subsequently,

$$\mathbb{D}\Big(P_{\mathbf{Z},\text{PPM}}^{n,\ell_n}\|Q^{\mathbf{0}}\Big) + (\delta_n' + \lambda_n + 1 - \eta_n) \times \left(\log\frac{|\mathcal{Z}|}{\left(\sup_{z:Q_1(z)>0}\left|\log\frac{Q_1(z)}{Q_0(z)}\right|\right)^2} n + \log\frac{e}{\delta_n' + \lambda_n + 1 - \eta_n}\right) \tag{2.62}$$

$$\leqslant \delta - \frac{1}{\sqrt{n}} + (\delta_n' + \lambda_n + 1 - \eta_n) \times \left(\log\frac{|\mathcal{Z}|}{\left(\sup_{z:Q_1(z)>0}\left|\log\frac{Q_1(z)}{Q_0(z)}\right|\right)^2} n + \log\frac{e}{\delta_n' + \lambda_n + 1 - \eta_n}\right) \tag{2.63}$$

$$= \delta - \frac{1}{\sqrt{n}} + \frac{3}{n^2}\left(\log\frac{|\mathcal{Z}|}{\left(\sup_{z:Q_1(z)>0}\left|\log\frac{Q_1(z)}{Q_0(z)}\right|\right)^2} n + \log\frac{en^2}{3}\right), \tag{2.64}$$

which is less than δ for large enough n. Therefore, Theorem 10 yields the existence of an $(\eta_n|\mathcal{M}_n|, |\mathcal{K}_n|, \epsilon, \delta)_{\max}^d$ where

$$\log|\mathcal{M}_n| \geqslant \omega\mathbb{D}(P_1\|P_0) n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon) n^{\frac{1}{4}} + O(\log n) + O(1) \tag{2.65}$$

with the required key size

$$\log|\mathcal{K}_n| \leqslant \max(0, \omega(\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0))n^{\frac{1}{2}} + O(n^{\frac{1}{4}}\sqrt{\log n})) \tag{2.66}$$

□

Converse of Theorem 7. The converse follows from Theorem 11 and the following lemma.

Lemma 11. *We have* $\lambda_d(w,\delta) \geqslant \gamma$ *for* $w = \sqrt{\frac{2\delta}{\chi_2(\chi_2(Q_1\|Q_0)(1-\frac{1}{\sqrt{n}})}}(1-\gamma)n^{\frac{1}{2}}$

Proof. Let $P_{\mathbf{X}}$ be a distribution over $\mathcal{X}^n$ such that $\mathbb{D}\Big(W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}\|Q^{\mathbf{0}}\Big) \leqslant \delta$. By [13, Equation (96)], we have $\delta \geqslant n\mathbb{D}\big(W_{Z|X} \circ \overline{P}_X\|Q_0\big)$ where $\overline{P}_X \triangleq \frac{1}{n}\sum_{i=1} P_{X_i}$. Moreover, we have by [13, Equation (11)],

$$\mathbb{D}\big(W_{Z|X} \circ \overline{P}_X\|Q_0\big) \geqslant \frac{\overline{P}_X(1)^2}{2}\chi_2(Q_1\|Q_0) - O(\overline{P}_X(1)^3). \tag{2.67}$$

Thus, $\overline{P}_X(1) \leqslant \sqrt{\frac{2\delta}{n\chi_2(Q_1\|Q_0)(1-O(\overline{P}_X(1)))}} = O(\frac{1}{\sqrt{n}})$. This implies that

$$\overline{P}_X(1) \leqslant \sqrt{\frac{2\delta}{n\chi_2(Q_1\|Q_0)(1-O\left(\frac{1}{\sqrt{n}}\right))}}. \tag{2.68}$$

We hence have

$$\mathbb{P}_{P_{\mathbf{X}}}(\mathrm{wt}(\mathbf{x}) \leqslant w) \geqslant 1 - \frac{\mathbb{E}_{P_{\mathbf{X}}}(\mathrm{wt}(\mathbf{X}))}{w} \tag{2.69}$$

$$= 1 - \frac{n\overline{P}_X(1)}{w} \tag{2.70}$$

$$\geqslant 1 - \frac{n\sqrt{\frac{2\delta}{n\chi_2(Q_1\|Q_0)(1-O\left(\frac{1}{\sqrt{n}}\right))}}}{w}. \tag{2.71}$$

□

□

2.6.2 Covertness with Variational Distance

In this subsection, we prove Theorem 8.

Achievability of Theorem 8. Let ϵ, δ, Γ and ω be as specified in Theorem 8. Similar to the analysis for relative entropy, we use Theorem 10 with $P_{\mathbf{X}}^n \triangleq P_{\mathbf{X},\text{PPM}}^{n,\ell_n}$ where ℓ_n is chosen to ensure $\frac{1}{2}\left\| P_{\mathbf{Z},\text{PPM}}^{n,\ell_n} - Q^{\mathbf{0}} \right\|_1 \leqslant \delta - \frac{1}{\sqrt{n}}$. Specifically, we wish to choose ℓ_n such that

$$\frac{\ell_n}{2}\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}} \geqslant \Gamma - \frac{e^{\frac{\Gamma^2}{2}}\sqrt{2\pi}}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.72}$$

The term Γ is what we would expect by inspection of (2.45) to obtain a variational distance with first order approximately δ. As we shall see next, the penalty term $\frac{e^{\frac{\Gamma^2}{2}}\sqrt{2\pi}}{\sqrt{\ell_n}}$ is the result of a Taylor series of the $\mathbb{Q}$ function around Γ. Formally, we choose $\sqrt{\ell_n}$ as a solution of the cubic equation

$$x^3 - 2\left(\Gamma + O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{\frac{n}{\chi_2(Q_1\|Q_0)}}x + 2\sqrt{2\pi}e^{\frac{\Gamma^2}{2}}\sqrt{\frac{n}{\chi_2(Q_1\|Q_0)}} = 0, \tag{2.73}$$

which will be sufficient to ensure (2.72). For a cubic equation of the form $x^3 - px + q = 0$, the roots are known in closed algebraic form and one of them is

$$2\sqrt{\frac{-p}{3}}\cos\left(\frac{1}{3}\arccos\left(\frac{3q}{2p}\sqrt{\frac{-3}{p}}\right)\right). \tag{2.74}$$

Thus,

$$\begin{aligned}\sqrt{\ell_n} = 2\sqrt{\frac{2\left(\Gamma + O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}} \\ \times \cos\left(\frac{1}{3}\arccos\left(-\frac{3\sqrt{2\pi}e^{\frac{\Gamma^2}{2}}}{2\left(\Gamma + O\left(\frac{1}{\sqrt{n}}\right)\right)}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{2\left(\Gamma + O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}}\right)\right)\end{aligned} \tag{2.75}$$

$$\overset{(a)}{=} 2\sqrt{\frac{2\left(\Gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}} \times \cos\left(\frac{1}{3}\left(\frac{\pi}{2}+\frac{3\sqrt{2\pi}e^{\frac{\Gamma^2}{2}}}{2\Gamma}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{2\Gamma\sqrt{n}}}+O\left(\frac{1}{\sqrt{n}}\right)\right)\right) \tag{2.76}$$

$$\overset{(b)}{=} 2\sqrt{\frac{2\left(\Gamma+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}}\left(\frac{\sqrt{3}}{2}-\frac{1}{2}\frac{\sqrt{2\pi}e^{\frac{\Gamma^2}{2}}}{2\Gamma}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{2\Gamma\sqrt{n}}}+O\left(\frac{1}{\sqrt{n}}\right)\right) \tag{2.77}$$

$$= \sqrt{\frac{2\Gamma\sqrt{n}}{\sqrt{\chi_2(Q_1\|Q_0)}}}-\sqrt{\frac{\pi}{2}}\frac{e^{\frac{\Gamma^2}{2}}}{\Gamma}+O\left(\frac{1}{n^{\frac{1}{4}}}\right), \tag{2.78}$$

where (a) and (b) follow since, for x close to zero, $\arccos(x) = \frac{\pi}{2} - x + O(x^2)$ and $\cos\left(\frac{\pi}{6}+x\right) = \frac{\sqrt{3}}{2} - \frac{1}{2}x + O(x^2)$, respectively. Consequently, upon choosing

$$\ell_n = 2\Gamma\sqrt{\frac{n}{\chi_2(Q_1\|Q_0)}}-\frac{2\sqrt{\pi}e^{\frac{\Gamma^2}{2}}n^{\frac{1}{4}}}{\sqrt{\Gamma}\chi_2(Q_1\|Q_0)^{\frac{1}{4}}}+O\left(1\right), \tag{2.79}$$

we satisfy (2.72). Combining with (2.45), we have

$$\frac{1}{2}\left\|P^{n,\ell_n}_{\mathbf{Z},\text{PPM}}-Q^{\mathbf{0}}\right\|_1 \leqslant 1-2\mathbb{Q}\left(\frac{\ell_n}{2}\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}}\right)+\frac{2}{\sqrt{\ell_n}}+O\left(\frac{1}{\sqrt{n}}\right) \tag{2.80}$$

$$\leqslant 1-2\mathbb{Q}\left(\Gamma-\frac{e^{\frac{\Gamma^2}{2}}\sqrt{2\pi}}{\sqrt{\ell_n}}+O\left(\frac{1}{\sqrt{n}}\right)\right)+\frac{2}{\sqrt{\ell_n}}+O\left(\frac{1}{\sqrt{n}}\right) \tag{2.81}$$

$$\overset{(a)}{\leqslant} 1-2\mathbb{Q}(\Gamma)-\frac{2}{\sqrt{\ell_n}}+\frac{2}{\sqrt{\ell_n}}+O\left(\frac{1}{\sqrt{n}}\right) \tag{2.82}$$

$$= \delta+O\left(\frac{1}{\sqrt{n}}\right), \tag{2.83}$$

where (a) follows from $Q(x-y) \geqslant Q(x) + \frac{e^{-\frac{x^2}{2}}y}{\sqrt{2\pi}}$ for $x \geqslant y \geqslant 0$. We can also control the term $O\left(\frac{1}{\sqrt{n}}\right)$ in (2.72) to guarantee that (2.83) is less than $\delta - \frac{1}{\sqrt{n}}$. The rest of the proof is exactly the same as the proof of Theorem 7 with the same parameter as in Table 2.6.1

except for ℓ_n. [2] $\square$

Converse for Theorem 8. The converse follows from Theorem 11 and the following two lemmas.

Lemma 12. *Assume that n is large enough and w is such that w/n and $w^2/n^{3/2}$ are small enough. We have*

$$\beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}} : \mathbf{x} \in \mathcal{X}^n, wt(\mathbf{x}) \geqslant w\}) \leqslant 1 - \mathbb{Q}\left(\mathbb{Q}^{-1}(\alpha) - \frac{w\sqrt{\chi_2(Q_1\|Q_0)}}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}} + \frac{w}{n} + \frac{w^2}{n^{\frac{3}{2}}}\right) \quad (2.84)$$

Proof. See Appendix 2.E. $\square$

We next establish a lower bound on $\lambda_d(w, \delta)$.

Lemma 13. *There exists $C > 0$ depending on Q_1 and Q_0 such that for all $\gamma \in [0, 1]$ and*

$$w = \frac{2\sqrt{n}}{\sqrt{\chi_2(Q_1\|Q_0)}}\mathbb{Q}^{-1}\left(\frac{1-\delta}{2} - \frac{C}{\sqrt{n}} - \gamma\right) \quad (2.85)$$

we have $\lambda_d(w, \delta) \geqslant \gamma$.

Proof. Let $P_{\mathbf{X}}$ be a PMF over $\mathcal{X}^n$ such that $\frac{1}{2}\left\|W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} - Q^{\mathbf{0}}\right\| \leqslant \delta$. We want to show that $p \triangleq \mathbb{P}_{P_{\mathbf{X}}}(\mathrm{wt}(\mathbf{X}) \leqslant w) \geqslant \gamma$ where w is defined in (2.85). By law of total probability, we can write

$$P_{\mathbf{X}} = pP_{\mathbf{X}}^{(1)} + (1-p)P_{\mathbf{X}}^{(2)} \quad (2.86)$$

such that $P_{\mathbf{X}}^{(1)}$ and $P_{\mathbf{X}}^{(2)}$ are two PMFs over $\mathcal{X}^n$ with $\min_{\mathbf{x}\in\mathrm{support}\left(P_{\mathbf{X}}^{(2)}\right)} \mathrm{wt}(\mathbf{x}) > w$. We

[2] By triangle inequality $\left|\left\|P_{\mathbf{Z}} - Q^{\mathbf{0}}\right\|_1 - \left\|Q_{\mathbf{Z}} - Q^{\mathbf{0}}\right\|_1\right| \leqslant \|P_{\mathbf{Z}} - Q_{\mathbf{Z}}\|_1$. This bound is tighter than what we establish for the relative entropy in Lemma 10. We can therefore use the same continuity argument for the total variation distance.

therefore have

$$2\delta \geqslant \left\| W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} - Q^{\mathbf{0}} \right\|_1 \tag{2.87}$$

$$\overset{(a)}{\geqslant} (1-p)\left\| W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}^{(2)} - Q^{\mathbf{0}} \right\|_1 - p\left\| W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}^{(1)} - Q^{\mathbf{0}} \right\|_1 \tag{2.88}$$

$$\geqslant (1-p)\left\| W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}^{(2)} - Q^{\mathbf{0}} \right\|_1 - 2p, \tag{2.89}$$

where (a) follows from the triangle inequality and (2.86). Furthermore, note that for all $\alpha \in]0,1[$,

$$\frac{1}{2}\left\| W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}^{(2)} - Q^{\mathbf{0}} \right\|_1 \geqslant 1 - \alpha - \beta_\alpha(Q^{\mathbf{0}}, W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}^{(2)}) \tag{2.90}$$

$$\geqslant 1 - \alpha - \beta_\alpha(Q^{\mathbf{0}}, \left\{ Q^{\mathbf{x}} : \mathbf{x} \in \text{support}\left(P_{\mathbf{X}}^{(2)}\right) \right\}) \tag{2.91}$$

$$\geqslant 1 - \alpha - \beta_\alpha(Q^{\mathbf{0}}, \{ Q^{\mathbf{x}} : \text{wt}(x) \geqslant w \}) \tag{2.92}$$

$$\overset{(a)}{\geqslant} -\alpha + \mathbb{Q}\left(\mathbb{Q}^{-1}(\alpha) - \frac{w\sqrt{\chi_2(Q_1\|Q_0)}}{\sqrt{n}} \right) - O\left(\frac{1}{\sqrt{n}}\right) \tag{2.93}$$

where (a) follows form Lemma 12 and $w = O(\sqrt{n})$. Setting $\alpha = \mathbb{Q}\left(\frac{w\sqrt{\chi_2(Q_1\|Q_0)}}{2\sqrt{n}} \right)$, and substituting the bound in (2.93) in (2.89), we obtain that

$$2\delta \geqslant (1-p)\left(1 - 2\mathbb{Q}\left(\frac{w\sqrt{\chi_2(Q_1\|Q_0}}{2\sqrt{n}} \right) \right) - 2p - O\left(\frac{1}{\sqrt{n}}\right) \tag{2.94}$$

$$\overset{(a)}{=} 2(1-p)\left(1 - 2\mathbb{Q}\left(\frac{\left(\frac{2\sqrt{n}}{\sqrt{\chi_2(Q_1\|Q_0)}} \mathbb{Q}^{-1}\left(\frac{1-\delta}{2} - \frac{C}{\sqrt{n}} - \gamma \right) \right) \sqrt{\chi_2(Q_1\|Q_0}}{2\sqrt{n}} \right) \right)$$

$$- 2p - O\left(\frac{1}{\sqrt{n}}\right) \tag{2.95}$$

$$= 2(1-p)\left(\delta + 2\gamma + \frac{C}{\sqrt{n}} \right) - 2p - O(\frac{1}{\sqrt{n}}) \tag{2.96}$$

$$\overset{(b)}{\geqslant} 2(1-p)(\delta + 2\gamma) - 2p \tag{2.97}$$

$$\geqslant 2(\delta + 2\gamma) - 4p, \tag{2.98}$$

where (a) follows from the definition of w, and (b) holds upon a proper choice of C. Thus, we have $p \geqslant \gamma$ as desired. □

□

2.6.3 Covertness with Probability of Missed Detection at Fixed Significance Level

In this subsection, we prove Theorem 9, in which $d(P,Q) = 1-\alpha-\beta_\alpha(Q,P)$ for $\alpha \in]0,1[$.

Lemma 14. *For two distributions $P_{\mathbf{Z}}$ and $Q_{\mathbf{Z}}$ over $\mathcal{Z}^n$, we have*

$$\left|d(P_{\mathbf{Z}},Q^{\mathbf{0}}) - d(Q_{\mathbf{Z}},Q^{\mathbf{0}})\right| \leqslant \frac{1}{2}\|P_{\mathbf{Z}} - Q_{\mathbf{Z}}\|_1. \tag{2.99}$$

Proof. Let $\mathcal{T} \subset \mathcal{Z}^n$ be such that $Q^{\mathbf{0}}(\mathcal{T}) \leqslant \alpha$ and $P_{\mathbf{Z}}(\mathcal{T}^c) = \beta_\alpha(Q^{\mathbf{0}},P_{\mathbf{Z}}) = 1-\alpha-d(P_{\mathbf{Z}},Q^{\mathbf{0}})$. We then have

$$\begin{aligned} d(Q_{\mathbf{Z}},Q^{\mathbf{0}}) &= 1-\alpha-\beta_\alpha(Q^{\mathbf{0}},Q_{\mathbf{Z}}) && (2.100)\\ &\geqslant 1-\alpha-Q_{\mathbf{Z}}(\mathcal{T}^c) && (2.101)\\ &\geqslant 1-\alpha-P_{\mathbf{Z}}(\mathcal{T}^c)-\frac{1}{2}\|P_{\mathbf{Z}}-Q_{\mathbf{Z}}\|_1. && (2.102) \end{aligned}$$

□

Achievability of Theorem 9. Fix $\epsilon \in]0,1[$, $\delta \in]0,1-\alpha[$, and n; we just show how we can choose ℓ_n such that $\beta_\alpha(Q^{\mathbf{0}}P^{n,\ell_n}_{\mathbf{Z},\mathrm{PPM}}) \geqslant 1-\alpha-\delta+\frac{1}{\sqrt{n}}$ and the rest of the proof is similar to the proof for variational distance. Let $\Lambda \triangleq Q^{-1}(1-\alpha-\delta)$ and $\Upsilon \triangleq Q^{-1}(\alpha)$; we wish to choose ℓ_n to satisfy

$$\ell_n\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}} - Q^{-1}\left(\alpha+\frac{1}{\sqrt{\ell_n}}\right) \leqslant \Lambda - \frac{\sqrt{2\pi}e^{\frac{\Lambda^2}{2}}}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.103}$$

Using $Q^{-1}(\alpha+x) = Q^{-1}(\alpha) - \sqrt{2\pi}e^{\frac{\left(Q^{-1}(\alpha)\right)^2}{2}}x + O(x^2) = \Upsilon + \sqrt{2\pi}e^{\frac{\Upsilon^2}{2}}x + O(x^2)$ for

x close to zero, the above inequality is equivalent to

$$\ell_n\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}}-\Upsilon+\frac{\sqrt{2\pi}e^{\frac{\Upsilon^2}{2}}}{\sqrt{\ell_n}}\leqslant\Lambda-\frac{\sqrt{2\pi}e^{\frac{\Lambda^2}{2}}}{\sqrt{\ell_n}}+O\left(\frac{1}{\sqrt{n}}\right). \tag{2.104}$$

We form the cubic equation

$$x^3-\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{\frac{n}{\chi_2(Q_1\|Q_0)}}x+\sqrt{2\pi}\left(e^{\frac{\Lambda^2}{2}}+e^{\frac{\Upsilon^2}{2}}\right)\sqrt{\frac{n}{\chi_2(Q_1\|Q_0)}}=0, \tag{2.105}$$

and set $\sqrt{\ell_n}$ to be its solution; Thus, similar to (2.78), we obtain

$$\begin{aligned}\sqrt{\ell_n}&=2\sqrt{\frac{\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}}\\&\quad\times\cos\left(\frac{1}{3}\arccos\left(\frac{3\sqrt{2\pi}\left(e^{\frac{\Lambda^2}{2}}+e^{\frac{\Upsilon^2}{2}}\right)}{-2\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}}\right)\right) &(2.106)\\&=2\sqrt{\frac{\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}}\\&\quad\times\cos\left(\frac{1}{3}\left(\frac{\pi}{2}+\frac{3\sqrt{2\pi}\left(e^{\frac{\Lambda^2}{2}}+e^{\frac{\Upsilon^2}{2}}\right)}{2\left(\Lambda+\Upsilon\right)}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{\left(\Lambda+\Upsilon\right)\sqrt{n}}}+O\left(\frac{1}{\sqrt{n}}\right)\right)\right) &(2.107)\\&=2\sqrt{\frac{\left(\Lambda+\Upsilon+O\left(\frac{1}{\sqrt{n}}\right)\right)\sqrt{n}}{3\sqrt{\chi_2(Q_1\|Q_0)}}}\\&\quad\times\left(\cos\left(\frac{\pi}{6}\right)-\sin\left(\frac{\pi}{6}\right)\frac{\sqrt{2\pi}\left(e^{\frac{\Lambda^2}{2}}+e^{\frac{\Upsilon^2}{2}}\right)}{2\left(\Lambda+\Upsilon\right)}\sqrt{\frac{3\sqrt{\chi_2(Q_1\|Q_0)}}{\left(\Lambda+\Upsilon\right)\sqrt{n}}}+O\left(\frac{1}{\sqrt{n}}\right)\right) &(2.108)\\&=\sqrt{\frac{(\Lambda+\Upsilon)\sqrt{n}}{\sqrt{\chi_2(Q_1\|Q_0)}}}-\sqrt{\frac{\pi}{2}}\frac{e^{\frac{\Lambda^2}{2}}+e^{\frac{\Upsilon^2}{2}}}{\Lambda+\Upsilon}+O\left(\frac{1}{\sqrt{n}}\right). &(2.109)\end{aligned}$$

Hence, if we choose

$$\ell_n = \frac{(\Lambda+\Upsilon)}{\sqrt{\chi_2(Q_1\|Q_0)}} n^{\frac{1}{2}} - \frac{\sqrt{2\pi}\left(e^{\frac{\Lambda^2}{2}} + e^{\frac{\Upsilon^2}{2}}\right)}{\sqrt{\Lambda+\Upsilon}\chi_2(Q_1\|Q_0)^{\frac{1}{4}}} n^{\frac{1}{4}} + O(1), \tag{2.110}$$

we have

$$\beta_\alpha(Q^{\mathbf{0}}, P^{n,\ell_n}_{\mathbf{Z},\text{PPM}}) \overset{(a)}{\geqslant} \mathbb{Q}\left(\ell_n \sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}} - Q^{-1}\left(\alpha + \frac{1}{\ell_n}\right)\right) - \frac{1}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{2.111}$$

$$\overset{(b)}{\geqslant} \mathbb{Q}\left(\Lambda - \frac{\sqrt{2\pi}e^{\frac{\Lambda^2}{2}}}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right)\right) - \frac{1}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{2.112}$$

$$\overset{(c)}{\geqslant} \mathbb{Q}\left(\Lambda\right) + \frac{1}{\sqrt{\ell_n}} - \frac{1}{\sqrt{\ell_n}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{2.113}$$

$$= 1 - \alpha - \delta + O\left(\frac{1}{\sqrt{n}}\right), \tag{2.114}$$

where (a) follows from (2.46), (b) follows from (2.103), and (c) follows from $\mathbb{Q}(\Lambda + x) = \mathbb{Q}(\Lambda) + \frac{x}{\sqrt{2\pi}e^{\frac{\Lambda^2}{2}}} + O(x^2)$ for x close to zero. □

The proof of the converse of Theorem 9 requires the following steps similar to those of covertness with variational distance. We lower-bound $\lambda_d(w,\delta)$ and then use Theorem 11.

Lemma 15. *There exists $C > 0$ depending on Q_1 and Q_0 such that for all $\gamma > 0$ and*

$$w = \frac{1}{\sqrt{\chi_2(Q_1\|Q_0)}}\left(\mathbb{Q}^{-1}\left(1 - \alpha - \delta - \frac{C}{\sqrt{n}} - \gamma\right) + \mathbb{Q}^{-1}(\alpha)\right), \tag{2.115}$$

we have $\lambda_d(w,\delta) \geqslant \gamma$.

Proof. It follows from the same steps as in the proof of Lemma 13. □

2.7 Conclusion

We have developed an approach to study covert communication when covertness is measured with several distances. It is legitimate to ask which metric would make the most sense from an operational perspective. While relative entropy is amenable to a fairly extensive information-theoretic analysis, as illustrated by our complete characterization of second-order asymptotics in Theorem 7, variational distance, and probability of false alarm are probably more adequate since they directly relate to the operation of an adversary attempting to detect the communication. Since measuring covertness in terms of variational distance does not impose any constraint on where the adversary operates on its ROC curve, we believe that variational distance is perhaps the most operationally relevant covertness metric.

Our results have been restricted to binary-input DMCs to obtain simple closed-form expressions but extensions to arbitrary finite input alphabets may be obtained following the approach of [12]; however several research questions remain open. We have not characterized the exact second-order asymptotics of covert communication with variational distance and probability of missed detection metrics; we conjecture, however, that the upper-bounds are achievable. We have also not characterized the second-order asymptotics for the number of key bits required; a close inspection of our current proof technique shows that we explicitly rely on a law of large numbers to analyze the number of key bits, which one would have to circumvent.

APPENDIX

2.A The Effect of Average Probability of Error on First Order Asymptotics

Assume that the optimal second order asymptotics are of the form

$$\log M^*(\epsilon, \delta) = f(\delta)\sqrt{n} + o(\sqrt{n}) \tag{2.116}$$

for some function f that is *strictly* increasing in δ and does not depend on ϵ, which is what we would hope to establish based on our results maximum probability of error. Now for $1 > \epsilon > 0$ and $\delta > 0$, pick $\frac{\epsilon}{1-\epsilon} > \alpha > 0$ and set $\epsilon' = (1+\alpha)\epsilon - \alpha$, and $\delta' = (1+\alpha)\delta$. Since $\delta' > \delta$ and because of (2.116), we expect for all blocklengths large enough that

$$\log M^*(\epsilon', \delta') > \log M^*(\epsilon, \delta).$$

Consider now a code with $M^*(\epsilon', \delta')$ codewords, and let $\widehat{Q'}^n$ be the distribution induced at the eavesdropper's output when randomizing uniformly over all codewords. Let $\mathcal{S}$ be a set (with repetition) of $N \triangleq \alpha M^*(\epsilon', \delta')$ all-zero codewords. We construct a new code by adding the set $\mathcal{S}$ to the previous code. The resulting code has average probability of error

$$\begin{aligned} P_{\text{avg}} &= \mathbb{P}\Big(\widehat{W} \neq W | W \in \mathcal{S}\Big)\mathbb{P}(W \in \mathcal{S}) + \mathbb{P}\Big(\widehat{W} \neq W | W \notin \mathcal{S}\Big)\mathbb{P}(W \notin \mathcal{S}) \\ &\leqslant \frac{M^*(\epsilon', \delta')}{M^*(\epsilon', \delta') + N}\epsilon' + \frac{N}{M^*(\epsilon', \delta') + N} = \frac{1}{1+\alpha}\epsilon' + \frac{\alpha}{1+\alpha} = \epsilon. \end{aligned}$$

It also induces a new distribution $\widehat{Q}^n$ at the eavesdropper's output when randomizing uniformly over all codewords, such that

$$\begin{aligned}\mathbb{D}\Big(\widehat{Q'}^n \| Q_0^{\otimes n}\Big) &= \mathbb{D}\bigg(\frac{M^*(\epsilon',\delta')}{M^*(\epsilon',\delta')+N}\widehat{Q}^n + \frac{N}{M^*(\epsilon',\delta')+N}Q_0^{\otimes n}\|Q_0^{\otimes n}\bigg)\\ &\leqslant \frac{M^*(\epsilon',\delta')}{M^*(\epsilon',\delta')+N}\delta' = \frac{1}{1+\alpha}\delta' = \delta.\end{aligned}$$

We therefore obtain a new code with probability of error less than ϵ and relative entropy less than δ, but with a number of codewords $M^*(\epsilon',\delta')+N$ that *exceeds* $M^*(\epsilon,\delta)$. This contradiction shows that (2.116) does not hold, so that either $f(\delta)$ is non-increasing or δ or depends on ϵ. In both cases, this suggests a result significantly more complex than what we were trying to establish. Note that this does not contradict existing results on covert capacity [12, 13] because these works have all been in the regime $\lim_{n\to\infty}\epsilon = 0$. Finally, from this example, we conclude that no strong converse exists when measuring reliability with an *average* probability of error and even the first-order asymptotics depend on a non-vanishing average probability of error. Hence, one could expect a much more challenging analysis of second-order asymptotics under the average probability of error constraint.

2.B Proof of Theorem 10

We first prove a channel coding result. Let $f : \mathcal{M} \to \mathcal{X}^n$ be an encoder. We define for the optimal decoder $P_{\text{avg}}^{\text{err}}(f)$ and $P_{\max}^{\text{err}}(f)$ as the average and maximum probability of error, respectively.

Lemma 16. *Let $P_{\mathbf{X}}$ be a probability distribution over $\mathcal{X}^n$ and $F : \mathcal{M} \to \mathcal{X}^n$ be a random encoder whose codewords are sample independently according to $P_{\mathbf{X}}$. Let $\mathcal{M}$, $\epsilon \in]0,1[$, and $\epsilon' \in]0,\epsilon[$ be such that*

$$\log|\mathcal{M}| \leqslant \mathbb{D}_s^{\epsilon-\epsilon'}\left(P_{\mathbf{X}}\otimes W_{Y|X}^{\otimes n}\|P_{\mathbf{X}}\otimes P^{\mathbf{0}}\right) + \log\frac{\epsilon'}{\chi_2\left(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}}\right)+1}. \tag{2.117}$$

We then have

$$\mathbb{E}_F\big(P_{\text{avg}}^{\text{err}}(F)\big) \leqslant \epsilon, \tag{2.118}$$

Additional, let $\mathcal{M}$, $\epsilon, \epsilon', \eta, p$ *be such that*

$$\log|\mathcal{M}| \leqslant \inf_{x\in\text{support}(P_{\mathbf{X}})} \mathbb{D}_s^{\epsilon-\eta}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{(1-\lambda_1)(1-\lambda_2)\eta}{\chi_2\left(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}}\right)+1}. \tag{2.119}$$

We then have

$$\mathbb{P}_F\Big(\exists\,\widetilde{\mathcal{M}}\subset\mathcal{M} : |\widetilde{\mathcal{M}}| \geqslant \eta\,|\mathcal{M}|\,, P_{\max}^{\text{err}}(F|_{\widetilde{\mathcal{M}}}) \leqslant \epsilon\Big) \geqslant p. \tag{2.120}$$

Proof. Note that (2.118) is a minor modification of [13] Using the bound in [13, Lemma 3], we have for any γ that

$$\mathbb{E}_F\big(P_{avg}^{err}(F)\big) \leqslant \mathbb{P}_{P_{\mathbf{X}}\otimes W_{Y|X}^{\otimes n}}\left(\log\frac{W_{Y|X}^{\otimes n}(\mathbf{Y}|\mathbf{X})}{P^{\mathbf{0}}(\mathbf{Y})} < \gamma\right) + \frac{|\mathcal{M}|}{\gamma}(\chi_2(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}})+1) \tag{2.121}$$

$$\leqslant \epsilon - \epsilon' + \frac{|\mathcal{M}|\,(\chi_2(W_{Y|X}^{\otimes n}\circ P_{\mathbf{X}}\|P^{\mathbf{0}})+1)}{\mathbb{D}_s^{\epsilon-\epsilon'}\left(P_{\mathbf{X}}\otimes W_{Y|X}^{\otimes n}\|P_{\mathbf{X}}\otimes P^{\mathbf{0}}\right)} \tag{2.122}$$

$$\leqslant \epsilon. \tag{2.123}$$

The proof of (2.120) is more subtle. For a fixed encoder f, let us consider the following sub-optimal decoding rule. Upon observing a sequence $\mathbf{y}\in\mathcal{Y}^n$, the receiver estimate the message as $\widehat{m}\in\mathcal{M}$ such that

$$\log\frac{W_{Y|X}^{\otimes n}(\mathbf{y}|f(\widehat{m}))}{P^{\mathbf{0}}(\mathbf{y})} \geqslant \gamma. \tag{2.124}$$

If there is no or there exists multiple such $\widehat{m}$, the decoder declares an error. When the

message is m and the receiver does not correctly decode, at least one of the two following events should happen.

$$\mathcal{E}_m^{(1)} \triangleq \left\{ \log \frac{W_{Y|X}^{\otimes n}(\mathbf{y}|f(m))}{P^{\mathbf{0}}(\mathbf{y})} < \gamma \right\} \tag{2.125}$$

$$\mathcal{E}_m^{(2)} \triangleq \left\{ \exists m' \in \mathcal{M} \setminus \{m\} : \log \frac{W_{Y|X}^{\otimes n}(\mathbf{y}|f(m'))}{P^{\mathbf{0}}(\mathbf{y})} \geqslant \gamma \right\} \tag{2.126}$$

Let $\epsilon_m^{(1)}$ and $\epsilon_m^{(2)}$ be the probabilities of $\mathcal{E}_m^{(1)}$ and $\mathcal{E}_m^{(2)}$, respectively, conditioned to the transmission of $f(m)$. Note that $\epsilon_m^{(1)}$ and $\epsilon_m^{(2)}$ are random variables for the random encoder F. By definition, we almost surely have

$$\epsilon_m^{(1)} \leqslant \sup_{\mathbf{x} \in \mathrm{support}(P_{\mathbf{X}})} \mathbb{P}_{P^{\mathbf{x}}} \left(\log \frac{W_{Y|X}^{\otimes n}(\mathbf{Y}|f(m))}{P^{\mathbf{0}}(\mathbf{Y})} < \gamma \right) \tag{2.127}$$

$$\overset{(a)}{\leqslant} \epsilon - \epsilon', \tag{2.128}$$

where (a) holds for $\gamma = \sup_{\mathbf{x} \in \mathrm{support}(P_{\mathbf{X}})} \mathbb{D}_s^{\epsilon-\epsilon'}\left(P^{\mathbf{x}} \| P^{\mathbf{0}}\right)$ and

$$\mathbb{E}_F \left(\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \epsilon_m^{(2)} \right) \leqslant \frac{|\mathcal{M}|}{e^{\gamma}} (\chi_2(W_{Y|X}^{\otimes n} \circ P_{\mathbf{X}} \| P^{\mathbf{0}}) + 1) \tag{2.129}$$

$$\overset{(a)}{\leqslant} (1 - \lambda_1)(1 - \lambda_2)\epsilon', \tag{2.130}$$

where (a) follows from the definition of $|\mathcal{M}|$ and γ. Note next that

$$\mathbb{P}_F\Big(\exists\, \widetilde{\mathcal{M}} \subset \mathcal{M} : |\widetilde{\mathcal{M}}| \geqslant \eta\,|\mathcal{M}|\,, P_{\max}^{\mathrm{err}}(F|_{\widetilde{\mathcal{M}}}) \leqslant \epsilon\Big) \tag{2.131}$$

$$\geqslant \mathbb{P}_F\left(\exists\, \widetilde{\mathcal{M}} \subset \mathcal{M} : |\widetilde{\mathcal{M}}| \geqslant \eta\,|\mathcal{M}|\,, \max_{m\in\widetilde{\mathcal{M}}} \epsilon_m^{(1)} \leqslant \epsilon - \epsilon', \max_{m\in\widetilde{\mathcal{M}}} \epsilon_m^{(2)} \leqslant \epsilon'\right) \tag{2.132}$$

$$\overset{(a)}{=} \mathbb{P}_F\left(\exists\, \widetilde{\mathcal{M}} \subset \mathcal{M} : |\widetilde{\mathcal{M}}| \geqslant \eta\,|\mathcal{M}|\,, \max_{m\in\widetilde{\mathcal{M}}} \epsilon_m^{(2)} \leqslant \epsilon'\right) \tag{2.133}$$

$$\overset{(b)}{\geqslant} \mathbb{P}_F\left(\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \epsilon_m^{(2)} \leqslant (1-\eta)\epsilon'\right) \tag{2.134}$$

$$\overset{(c)}{\geqslant} 1 - \frac{\mathbb{E}_F\left(\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \epsilon_m^{(2)}\right)}{(1-\eta)\epsilon'} \tag{2.135}$$

$$\overset{(d)}{\geqslant} p, \tag{2.136}$$

where (a) follows from (2.128), (b) and (c) follow from Markov's inequality, and (d) follows from (2.130).

□

By Lemma 16 and (2.25), with probability at least p (with respect to the random coding), for each $k \in K$, there exists a subset $\widetilde{\mathcal{M}}_k$ of $\mathcal{M}$ of size at least $\eta\,|\mathcal{M}|$ such that when the value of the key is k and the message is restricted to the set $\widetilde{\mathcal{M}}$ the maximum probability of error (of optimal decoder) is less than ϵ. By permuting the messages, we can assume without loss of generality that $\widetilde{\mathcal{M}}_k = \widetilde{\mathcal{M}}$ for all $k \in \mathcal{K}$. Since the performance of the subcodes corresponding to each value of the key is independent, with probability at least $p^{|\mathcal{K}|}$ (with respect to the random coding), there exists a message set $\widetilde{\mathcal{M}}$ of size at least $\eta\,|\mathcal{M}|$ and an encoder $\widetilde{F} : \widetilde{\mathcal{M}} \times \mathcal{K} \to \mathcal{X}^n$ such that the maximum probability under optimal decoding rule is at most ϵ. Therefore, by (1.33) and (2.27), there exists an encoder $\widetilde{f} : \widetilde{\mathcal{M}} \times \mathcal{K} \to \mathcal{X}^n$ such that the maximum probability of error under optimal decoding is less than ϵ, and its

codewords are also a subset of the codewords of an encoder $f : \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n$ with

$$\left\| \frac{1}{|\mathcal{M}|\,|\mathcal{K}|} \sum_{m\in\mathcal{M},k\in\mathcal{K}} Q^{f(m,k)} - W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} \right\|_1 < \delta + \lambda. \tag{2.137}$$

Note now that

$$d\left(\frac{1}{\left|\widetilde{\mathcal{M}}\right|\,|\mathcal{K}|} \sum_{m\in\widetilde{\mathcal{M}},k\in\mathcal{K}} Q^{\widetilde{f}(m,k)}, Q^{\mathbf{0}} \right) \tag{2.138}$$

$$\overset{(a)}{\leqslant} L\left(\left\| \frac{1}{\left|\widetilde{\mathcal{M}}\right|\,|\mathcal{K}|} \sum_{m\in\widetilde{\mathcal{M}},k\in\mathcal{K}} Q^{\widetilde{f}(m,k)} - W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} \right\|_1 \right) + d(W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}, Q^{\mathbf{0}}) \tag{2.139}$$

$$\overset{(b)}{\leqslant} L\left(\left\| \frac{1}{|\mathcal{M}|\,|\mathcal{K}|} \sum_{m\in\mathcal{M},k\in\mathcal{K}} Q^{f(m,k)} - W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}} \right\|_1 + 1 - \eta \right) + d(W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}, Q^{\mathbf{0}}) \tag{2.140}$$

$$\leqslant L(\delta + \lambda + 1 - \eta) + d(W_{Z|X}^{\otimes n} \circ P_{\mathbf{X}}, Q^{\mathbf{0}}), \tag{2.141}$$

where (a) follows since d is L-continuous at $Q^{\mathbf{0}}$, and (b) follows from the trainable inequality.

2.C Proof of Lemma 9

We first prove a technical lemma.

Lemma 17. *Suppose* $\mathbf{Z} \triangleq (Z_1, \cdots, Z_m)$ *is distributed according to* $Q_0^{\otimes m}$. *Define* $A(z) \triangleq \frac{Q_1(z)-Q_0(z)}{Q_0(z)}$ *and let* $B \triangleq \frac{1}{m}\sum_{i=1}^m A(Z_i)$. *Also define* $\underline{B}$ *with distribution* $\mathbb{P}(\underline{B} = b) \triangleq \mathbb{P}(B = b | B \neq -1)$ *and* $\underline{C} \triangleq \log(1 + \underline{B})$. *There exists a constant* $0 \leqslant \tau < 1$ *depending*

only on the channel such that

$$\mathbb{E}(\underline{B}) = O(\tau^m), \tag{2.142}$$

$$\mathbb{E}\left(\underline{B}^2\right) = \frac{\chi_2(Q_1\|Q_0)}{m} + O(\tau^m), \tag{2.143}$$

$$\mathbb{E}\left(\underline{B}^3\right) = O\left(\frac{1}{m^2}\right), \tag{2.144}$$

$$\mathbb{E}\left(\underline{B}^4\right) = \frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right), \tag{2.145}$$

and

$$\mathbb{E}(\underline{C}) = -\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right) \tag{2.146}$$

$$\text{Var}(\underline{C}) = \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right) \tag{2.147}$$

$$\mathbb{E}\left(|\underline{C} - \mathbb{E}(\underline{C})|^3\right) \leqslant \frac{\chi_2(Q_1\|Q_0)^{\frac{3}{2}}}{m^{\frac{3}{2}}} + O\left(\frac{1}{m^{\frac{9}{4}}}\right). \tag{2.148}$$

Suppose $\widetilde{\mathbf{Z}} \triangleq (\widetilde{Z}_1, \cdots, \widetilde{Z}_m)$ *is distributed according to* $P^{m,1}_{\mathbf{Z},\text{PPM}}$. *Define* $\widetilde{B} \triangleq \frac{1}{m}\sum_{i=1}^m A(\widetilde{Z}_i)$, $\underline{\widetilde{B}}$ *with distribution* $\mathbb{P}\left(\underline{\widetilde{B}} = b\right) \triangleq \mathbb{P}\left(\widetilde{B} = b | \widetilde{B} \neq -1\right)$ *and* $\underline{\widetilde{C}} \triangleq \log(1 + \underline{\widetilde{B}})$. *Then*

$$\mathbb{E}\left(\underline{\widetilde{B}}\right) = \frac{\chi_2(Q_1\|Q_0)}{m} + O(\tau^m), \tag{2.149}$$

$$\mathbb{E}\left(\underline{\widetilde{B}}^2\right) = \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right), \tag{2.150}$$

$$\mathbb{E}\left(\underline{\widetilde{B}}^3\right) = O\left(\frac{1}{m^2}\right), \tag{2.151}$$

$$\mathbb{E}\left(\underline{\widetilde{B}}^4\right) = \frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right), \tag{2.152}$$

and

$$\mathbb{E}\Big(\widetilde{\underline{C}}\Big) = \frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right) \tag{2.153}$$

$$\mathrm{Var}\Big(\widetilde{\underline{C}}\Big) = \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right) \tag{2.154}$$

$$\mathbb{E}\Big(|\widetilde{\underline{C}} - \mathbb{E}\Big(\widetilde{\underline{C}}\Big)|^3\Big) \leqslant \frac{\chi_2(Q_1\|Q_0)^{\frac{3}{2}}}{m^{\frac{3}{2}}} + O\left(\frac{1}{m^{\frac{9}{4}}}\right). \tag{2.155}$$

Proof. We first show that the moments of B and $\underline{B}$ are close. Define $\kappa \triangleq \max_{z:Q_0(z)>0} \left|\frac{Q_1(z)-Q_0(z)}{Q_0(z)}\right|$ and $\tau \triangleq \mathbb{P}_{Q_0}(A(Z) = -1)$, which is strictly less than one because $Q_1 \ll Q_0$, and therefore, supp$Q_1 \cap$ supp$Q_0 \neq \emptyset$. Notice that $\frac{1}{2}\|P_B - P_{\underline{B}}\|_1 = \tau^m$, $|B| \leqslant \kappa$ and $|\underline{B}| \leqslant \kappa$ almost surely and that κ does not depend on m. For any $q \geqslant 1$, we therefore obtain

$$|\mathbb{E}(B^q) - \mathbb{E}(\underline{B}^q)| \leqslant \sum_b |b^q\left(\mathbb{P}(B=b) - \mathbb{P}(\underline{B}=b)\right)| \tag{2.156}$$

$$\leqslant \frac{1}{2}\kappa^q \|P_B - P_{\underline{B}}\|_1 \tag{2.157}$$

$$= \kappa^q \tau^m. \tag{2.158}$$

Therefore, it is sufficient to find $\mathbb{E}(B^i)$ for $i = 1, 2, 3, 4$. Note that

$$\mathbb{E}_{Q_0}(A(Z)) = 0 \tag{2.159}$$

$$\mathbb{E}_{Q_1}(A(Z)) = \chi_2(Q_1\|Q_0) \tag{2.160}$$

$$\mathrm{Var}_{Q_0}(A(Z)) = \chi_2(Q_1\|Q_0). \tag{2.161}$$

Hence, we obtain

$$\mathbb{E}(B) = \mathbb{E}\left(\frac{1}{m}\sum_{i=1}^{m} A(Z_i)\right) = 0, \tag{2.162}$$

$$\mathbb{E}\big(B^2\big) = \frac{1}{m^2}\mathbb{E}\left(\left(\sum_{i=1}^{m} A(Z_i)\right)^2\right) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{E}(A(Z_i)A(Z_j)) = \frac{\chi_2(Q_1\|Q_0)}{m} \tag{2.163}$$

$$\mathbb{E}\big(B^3\big) = \frac{1}{m^3}\mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} A(Z_i)A(Z_k)A(Z_j)\right) = \frac{\sum_{i=1}^{m}\mathbb{E}(A(Z_i)^3)}{m^3} = O\left(\frac{1}{m^2}\right) \tag{2.164}$$

$$\mathbb{E}\big(B^4\big) = \frac{1}{m^4}\mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{t=1}^{m} A(Z_i)A(Z_j)A(Z_k)A(Z_t)\right) \tag{2.165}$$

$$= \frac{\sum_{i=1}^{m} A(Z_i)^4 + 2\sum_{i=1}^{m}\sum_{j=i+1}^{m}\mathbb{E}(A(Z_i)^2)\mathbb{E}(A(Z_j)^2)}{m^4} \tag{2.166}$$

$$= \frac{\chi_2(Q_1\|Q_0)^2}{m^2} + \frac{-\chi_2(Q_1\|Q_0) + \mathbb{E}_{Q_0}(A(Z)^4)}{m^3} = \frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right). \tag{2.167}$$

To find the expected value of $\underline{C}$, note that $\ln(1+x) \leqslant x - \frac{x^2}{2} + \frac{x^3}{3}$, and therefore, we have

$$\mathbb{E}(\underline{C}) = \mathbb{E}(\log(1+\underline{B})) \leqslant \mathbb{E}(\underline{B}) - \mathbb{E}\left(\frac{\underline{B}^2}{2}\right) + \mathbb{E}\left(\frac{\underline{B}^3}{3}\right) = -\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right). \tag{2.168}$$

For any $0 < a < 1$ and $x > -a$, we also have $x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4(1-a)^4} \leqslant \ln(1+x)$; thus,

$$\mathbb{E}(\underline{C}) \geqslant \mathbb{E}(\underline{B}) - \mathbb{E}\left(\frac{\underline{B}^2}{2}\right) + \mathbb{E}\left(\frac{\underline{B}^3}{3}\right) - \mathbb{E}\left(\frac{\underline{B}^4}{4(1+\min_{b:\mathbb{P}(\underline{B}=b)>0} b)^4}\right) \tag{2.169}$$

$$= -\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right) - \frac{\frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right)}{4(1+\min_{b:\mathbb{P}(\underline{B}=b)>0} b)^4} \tag{2.170}$$

$$= -\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right) - \frac{\frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right)}{4(1+\min_{z:Q_0(z)>0,Q_1(z)} A(z))^4} \tag{2.171}$$

$$\overset{(a)}{=} -\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right), \tag{2.172}$$

where (a) follows since $\min_{z:Q_0(z)>0,Q_1(z)>0} A(z) > -1$. Furthermore, for the variance of

$\underline{C}$, since $\log^2(1+x) \geqslant \left(x - \frac{x^2}{2}\right)^2 - x^4$, we have

$$\begin{aligned}
\text{Var}(\underline{C}) &= \mathbb{E}\big(\underline{C}^2\big) - \mathbb{E}(\underline{C})^2 && (2.173)\\
&\geqslant \mathbb{E}\left(\left(\underline{B} - \frac{\underline{B}^2}{2}\right)^2\right) - \mathbb{E}\big(\underline{B}^4\big) + O\left(\frac{1}{m^2}\right) && (2.174)\\
&= \mathbb{E}\big(\underline{B}^2\big) - \mathbb{E}\big(\underline{B}^3\big) - \mathbb{E}\left(\frac{3}{4}\underline{B}^4\right) + O\left(\frac{1}{m^2}\right) && (2.175)\\
&= \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right). && (2.176)
\end{aligned}$$

Moreover, for $-1 < a < 0$ and $x > -a$, we have $\log^2(1+x) \leqslant \left(x - \frac{x^2}{2} + \frac{x^3}{3(1-a)^3}\right)^2$. Therefore,

$$\begin{aligned}
\text{Var}(\underline{C}) &= \mathbb{E}\big(\underline{C}^2\big) - \mathbb{E}(\underline{C})^2 && (2.177)\\
&\geqslant \mathbb{E}\left(\left(\underline{B} - \frac{\underline{B}^2}{2} - \frac{\underline{B}^3}{3(1-\min_{z:Q_0(z)>0,Q_1(z)>0} A(z))^3}\right)\right) + O\left(\frac{1}{m^2}\right) && (2.178)\\
&= \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right). && (2.179)
\end{aligned}$$

Finally, for the third moment of C, note that

$$\begin{aligned}
\mathbb{E}\big(|\underline{C} - \mathbb{E}(\underline{C})|^3\big) &= \mathbb{E}\Big(\big((\underline{C} - \mathbb{E}(\underline{C}))^4\big)^{\frac{3}{4}}\Big) && (2.180)\\
&\overset{(a)}{\leqslant} \big(\mathbb{E}\big((\underline{C} - \mathbb{E}(\underline{C}))^4\big)\big)^{\frac{3}{4}} && (2.181)\\
&= \big(\mathbb{E}\big(\underline{C}^4 - 4\underline{C}^3\mathbb{E}(\underline{C}) + 6\underline{C}^2(\mathbb{E}(\underline{C}))^2 - 4C(\mathbb{E}(\underline{C}))^3 + (\mathbb{E}(\underline{C}))^4\big)\big)^{\frac{3}{4}} && (2.182)\\
&= \frac{\chi_2(Q_1\|Q_0)^{\frac{3}{2}}}{m^{\frac{3}{2}}} + O\left(\frac{1}{m^{\frac{9}{4}}}\right), && (2.183)
\end{aligned}$$

where (a) follows from Jensen's inequality.

We now calculate the moments with respect to PPM distribution. First note that if we define $P^i_{\mathbf{Z}}(\mathbf{z}) \triangleq Q_1(z_i)\prod_{j\neq i} Q_0(z_j)$, then we can write $P^{m,1}_{\mathbf{Z},\text{PPM}} = \frac{1}{m}P^i_{\mathbf{Z}}$; therefore, for any

function of $\widetilde{\mathbf{Z}}$ such as $f(\widetilde{\mathbf{Z}})$, we have $\mathbb{E}_{P^{m,1}_{\mathbf{Z},\mathrm{PPM}}}\Big(f(\widetilde{\mathbf{Z}})\Big) = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_{P^i_{\mathbf{Z}}}\Big(f(\widetilde{\mathbf{Z}})\Big)$. Moreover, if f is invariant with respect to all permutation of elements of $\mathbf{z}$, then $\mathbb{E}_{P^{m,1}_{\mathbf{Z},\mathrm{PPM}}}\Big(f(\widetilde{\mathbf{Z}})\Big) = \mathbb{E}_{P^1_{\mathbf{Z}}}\Big(f(\widetilde{\mathbf{Z}})\Big)$. One can check that this property holds for all moments of $\widetilde{B}$ and $\widetilde{C}$. Therefore, in the sequel, we assume that $\widetilde{\mathbf{Z}}$ is distributed according to $P^1_{\mathbf{Z}}$. Hence, we have

$$\mathbb{E}\Big(\widetilde{B}\Big) = \mathbb{E}\left(\frac{1}{m}\sum_{i=1}^{m}A(\widetilde{Z}_i)\right) = \frac{\chi_2(Q_1\|Q_0)}{m}, \tag{2.184}$$

$$\mathbb{E}\Big(\widetilde{B}^2\Big) = \frac{1}{m^2}\mathbb{E}\left(\left(\sum_{i=1}^{m}A(\widetilde{Z}_i)\right)^2\right) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{E}\Big(A(\widetilde{Z}_i)A(\widetilde{Z}_j)\Big) \tag{2.185}$$

$$= \frac{(m-1)\chi_2(Q_1\|Q_0) + \mathbb{E}_{Q_1}(A(Z)^2)}{m^2} = \frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right) \tag{2.186}$$

$$\mathbb{E}\Big(\widetilde{B}^3\Big) = \frac{1}{m^3}\mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}A(\widetilde{Z}_i)A(\widetilde{Z}_k)A(\widetilde{Z}_j)\right) \tag{2.187}$$

$$= \frac{\sum_{i=1}^{m}\mathbb{E}\Big(A(\widetilde{Z}_i)^3\Big) + \sum_{i=2}^{m}\mathbb{E}\Big(A(\widetilde{Z}_1)A(\widetilde{Z}_i)^2\Big)}{m^3} = O\left(\frac{1}{m^2}\right) \tag{2.188}$$

$$\mathbb{E}\Big(\widetilde{B}^4\Big) = \frac{1}{m^4}\mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{t=1}^{m}A(\widetilde{Z}_i)A(\widetilde{Z}_j)A(\widetilde{Z}_k)A(\widetilde{Z}_t)\right) \tag{2.189}$$

$$= \frac{1}{m^4}\left(\sum_{i=1}^{m}A(\widetilde{Z}_i)^4 + 2\sum_{i=1}^{m}\sum_{j=i+1}^{m}\mathbb{E}\Big(A(\widetilde{Z}_i)^2\Big)\mathbb{E}\Big(A(\widetilde{Z}_j)^2\Big) + \sum_{i=2}^{m}\mathbb{E}\Big(A(\widetilde{Z}_1)A(\widetilde{Z}_i)^3\Big)\right) \tag{2.190}$$

$$= \frac{\chi_2(Q_1\|Q_0)^2}{m^2} + O\left(\frac{1}{m^3}\right). \tag{2.191}$$

One can show that $\Big|\mathbb{E}(\underline{\widetilde{B}}^q) - \mathbb{E}\Big(\widetilde{B}^q\Big)\Big| = O(\tau^m)$ as was done earlier for B and $\underline{B}$. Using then the same bounds for $\log(1+x)$ as before and the above calculations, we obtain the desired bounds for the moments of $\widetilde{C}$. □

We are now ready to prove Lemma 9. By [46, Lemma 1], we have for n large enough

and some positive constant C_1, C_2, C_3, and C_4,

$$\mathbb{D}\Big(P^{n,\ell}_{\mathbf{Z},\text{PPM}}\|Q^{\mathbf{0}}\Big) \leqslant \frac{\ell}{2m}\chi_2(Q_1\|Q_0) + \frac{\ell C_1}{m^2} + \frac{\ell C_2}{m^3} \tag{2.192}$$

$$\leqslant \frac{\ell}{2m}\chi_2(Q_1\|Q_0) + \frac{\ell C_3}{m^2} \tag{2.193}$$

$$\leqslant \frac{\ell}{2(n/\ell-1)}\chi_2(Q_1\|Q_0) + \frac{\ell C_3}{(n/\ell-1)^2} \tag{2.194}$$

$$\leqslant \frac{\ell^2}{2n}\chi_2(Q_1\|Q_0) + \frac{\ell^3 C_4}{n^2}. \tag{2.195}$$

To prove (2.45), note that for any two distributions P and Q, one can write $\frac{1}{2}\|P-Q\|_1 = \mathbb{P}_P(P(X) \geqslant Q(X)) - \mathbb{P}_Q(P(X) \geqslant Q(X))$. Therefore, if $\mathbf{Z} = (Z_1, \cdots, Z_n)$ and $\widetilde{\mathbf{Z}} = (\widetilde{Z}_1, \cdots, \widetilde{Z}_n)$ are distributed according to $Q^{\mathbf{0}}$ and $P^{n,\ell}_{\mathbf{Z},\text{PPM}}$, respectively, then

$$\frac{1}{2}\Big\|P^{n,\ell}_{\mathbf{Z},\text{PPM}} - Q^{\mathbf{0}}\Big\|_1 = \mathbb{P}\Big(P^{n,\ell}_{\mathbf{Z},\text{PPM}}(\widetilde{\mathbf{Z}}) \geqslant Q^{\mathbf{0}}(\widetilde{\mathbf{Z}})\Big) - \mathbb{P}\Big(P^{n,\ell}_{\mathbf{Z},\text{PPM}}(\mathbf{Z}) \geqslant Q^{\mathbf{0}}(\mathbf{Z})\Big) \tag{2.196}$$

$$= \mathbb{P}\left(\log\frac{P^{n,\ell}_{\mathbf{Z},\text{PPM}}(\widetilde{\mathbf{Z}})}{Q^{\mathbf{0}}(\widetilde{\mathbf{Z}})} \geqslant 0\right) - \mathbb{P}\left(\log\frac{P^{n,\ell}_{\mathbf{Z},\text{PPM}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant 0\right). \tag{2.197}$$

However, note that $P^{n,\otimes\ell}_{\mathbf{Z},\text{PPM}} = (P^{m,1}_{\mathbf{Z},\text{PPM}})^{\ell} \otimes Q_0^{\otimes r}$, and for any $\mathbf{z} = (z_1, \cdots, z_n)$ with $\mathbf{z}_i = (z_{(i-1)m+1}, \cdots, z_{im})$, we have

$$\log\left(\frac{P^{n,\ell}_{\mathbf{Z},\text{PPM}}(\mathbf{z})}{Q^{\mathbf{0}}(\mathbf{z})}\right) = \sum_{i=1}^{\ell}\log\left(\frac{P^{m,1}_{\mathbf{Z},\text{PPM}}(\mathbf{z}_i)}{Q_0^{\otimes m}(\mathbf{z}_i)}\right) \tag{2.198}$$

$$= \sum_{i=1}^{\ell}\log\left(\frac{\sum_{j=1}^{m}\frac{1}{m}Q_1(z_{(i-1)m+j})\prod_{k\neq j}Q_0(z_{(i-1)m+k})}{Q_0^{\otimes m}(\mathbf{z}_i)}\right) \tag{2.199}$$

$$= \sum_{i=1}^{\ell}\log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{Q_1(z_{(i-1)m+j})}{Q_0(z_{(i-1)m+j})}\right) \tag{2.200}$$

$$= \sum_{i=1}^{\ell}\log\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{Q_1(z_{(i-1)m+j})-Q_0(z_{(i-1)m+j})}{Q_0(z_{(i-1)m+j})}\right) \tag{2.201}$$

$$= \sum_{i=1}^{\ell}\log\left(1+\frac{1}{m}\sum_{j=1}^{m}A(z_{(i-1)m+j})\right), \tag{2.202}$$

for $A(z) \triangleq \frac{Q_1(z)-Q_0(z)}{Q_0(z)}$. Thus, if we define

$$B_i \triangleq \frac{1}{m}\sum_{j=1}^{m} A(Z_{(i-1)m+j}) \tag{2.203}$$

$$C_i \triangleq \log(1+B_i) \tag{2.204}$$

$$\widetilde{B}_i \triangleq \frac{1}{m}\sum_{j=1}^{m} A(\widetilde{Z}_{(i-1)m+j}) \tag{2.205}$$

$$\widetilde{C}_i \triangleq \log(1+\widetilde{B}_i). \tag{2.206}$$

Then, we have

$$\mathbb{P}\left(\log\frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\widetilde{\mathbf{Z}})}{Q^{\mathbf{0}}(\widetilde{\mathbf{Z}})} \geqslant 0\right) - \mathbb{P}\left(\log\frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant 0\right) = \mathbb{P}\left(\sum_{i=1}^{\ell}\widetilde{C}_i \geqslant 0\right) - \mathbb{P}\left(\sum_{i=1}^{\ell} C_i \geqslant 0\right). \tag{2.207}$$

Furthermore, we consider two iid sequences $\underline{B}_1, \cdots, \underline{B}_\ell$ and $\underline{\widetilde{B}}_1, \cdots, \underline{\widetilde{B}}_\ell$ with distribution

$$\mathbb{P}(\underline{B}_i = b) \triangleq \mathbb{P}(B_i = b | B_i \neq -1) \tag{2.208}$$

$$\mathbb{P}\Big(\underline{\widetilde{B}}_i = b\Big) \triangleq \mathbb{P}\Big(\widetilde{B}_i = b | \widetilde{B}_i \neq -1\Big). \tag{2.209}$$

Note that with $\tau \triangleq \mathbb{P}(A(Z_1) = -1) < 1$, we have $\mathbb{P}(B_i = -1) = \tau^m$. By defining $\underline{C}_i \triangleq \log(1+\underline{B}_i)$ and $\underline{\widetilde{C}}_i \triangleq \log(1+\underline{\widetilde{B}}_i)$, we will deal with random variables that take finite values with probability one. We now replace $\widetilde{C}_i$ by $\underline{\widetilde{C}}_i$, and since $\underline{\widetilde{C}}_1, \cdots \underline{\widetilde{C}}_\ell$ are iid, we apply Theorem 6 in Chapter 1 to obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{\ell}\widetilde{C}_i \geqslant 0\right) &= \mathbb{P}\left(\sum_{i=1}^{\ell}\widetilde{C}_i \geqslant 0 \Big| \forall i, \widetilde{B}_i \neq -1\right)\mathbb{P}\Big(\forall i, \widetilde{B}_i \neq -1\Big) \\ &\qquad + \mathbb{P}\left(\sum_{i=1}^{\ell}\widetilde{C}_i \geqslant 0 \Big| \exists i : \widetilde{B}_i = -1\right)\mathbb{P}\Big(\exists i : \widetilde{B}_i = -1\Big) \qquad (2.210) \\ &\leqslant \mathbb{P}\left(\sum_{i=1}^{\ell}\widetilde{C}_i \geqslant 0 \Big| \forall i, \widetilde{B}_i \neq -1\right) + \ell\tau^m \qquad (2.211)\end{aligned}$$

$$= \mathbb{P}\left(\sum_{i=1}^{\ell} \widetilde{\underline{C}}_i \geqslant 0\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{2.212}$$

$$\leqslant \mathbb{Q}\left(\frac{-\sum_{i=1}^{\ell} \mathbb{E}\left(\widetilde{\underline{C}}_i\right)}{\sqrt{\ell \sum_{i=1}^{\ell} \mathrm{Var}\left(\widetilde{\underline{C}}_i\right)}}\right) + \frac{\sum_{i=1}^{\ell} \mathbb{E}\left(\left|\widetilde{\underline{C}}_i - \mathbb{E}\left(\widetilde{\underline{C}}_i\right)\right|^3\right)}{\left(\sum_{i=1}^{\ell} \mathrm{Var}\left(\widetilde{\underline{C}}_i\right)\right)^{\frac{3}{2}}} \tag{2.213}$$

$$\overset{(a)}{\leqslant} \mathbb{Q}\left(-\frac{\ell\left(\frac{\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{m^2}\right)\right)}{\sqrt{\ell\left(\frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right)\right)}}\right) + \frac{\ell\left(\frac{\chi_2(Q_1\|Q_0)^{\frac{3}{2}}}{m^{\frac{3}{2}}} + O\left(\frac{1}{m^{\frac{9}{4}}}\right)\right)}{\left(\ell\left(\frac{\chi_2(Q_1\|Q_0)}{m} + O\left(\frac{1}{m^2}\right)\right)\right)^{\frac{3}{2}}} \tag{2.214}$$

$$= \mathbb{Q}\left(-\frac{1}{2}\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}\right) + \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right), \tag{2.215}$$

where (a) follows from Lemma 17. Analogously, using Lemma 17 we obtain

$$\mathbb{P}\left(\sum_{i=1}^{\ell} C_i \geqslant 0\right) \geqslant \mathbb{Q}\left(\frac{1}{2}\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}\right) - \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right), \tag{2.216}$$

and therefore, we have (2.45).

To lower-bound $\beta_\alpha(Q^{\mathbf{0}}, P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}})$, we use the Neyman-Pearson lemma, which states that if $\alpha \leqslant \mathbb{P}\left(\log \frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant \gamma\right)$ for some γ, then we have

$$\beta_\alpha(Q^{\mathbf{0}}, P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}) \geqslant \mathbb{P}\left(\log \frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\widetilde{\mathbf{Z}})}{Q^{\mathbf{0}}(\widetilde{\mathbf{Z}})} \leqslant \gamma\right), \tag{2.217}$$

where $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$ are defined as before. Using Theorem 6 in Chapter 1 again, similar to (2.216), we obtain

$$\mathbb{P}\left(\log \frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant \gamma\right) \geqslant \mathbb{Q}\left(\frac{\gamma + \frac{\ell\chi_2(Q_1\|Q_0)}{2m}}{\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}}\right) - \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.218}$$

Accordingly, to ensure that $\mathbb{P}\left(\log \frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \leqslant \gamma\right) \geqslant \alpha$, we choose

$$\gamma = \sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}\mathbb{Q}^{-1}\left(\alpha + \frac{1}{\sqrt{\ell}}\right) - \frac{\ell\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.219}$$

Thus, we have

$$\beta_\alpha(Q^{\mathbf{0}}, P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}) \tag{2.220}$$

$$\geqslant \mathbb{P}\left(\log \frac{P^{n,\ell}_{\mathbf{Z},\mathrm{PPM}}(\widetilde{\mathbf{Z}})}{Q^{\mathbf{0}}(\widetilde{\mathbf{Z}})} \leqslant \sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}\mathbb{Q}^{-1}\left(\alpha + \frac{1}{\sqrt{\ell}}\right) - \frac{\ell\chi_2(Q_1\|Q_0)}{2m} + O\left(\frac{1}{\sqrt{n}}\right)\right) \tag{2.221}$$

$$\geqslant \mathbb{Q}\left(\frac{-\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}\mathbb{Q}^{-1}\left(\alpha + \frac{1}{\sqrt{\ell}}\right) + \frac{\ell\chi_2(Q_1\|Q_0)}{2m} + \frac{\ell\chi_2(Q_1\|Q_0)}{2m}}{\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}}}\right) - \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{2.222}$$

$$= \mathbb{Q}\left(\sqrt{\frac{\ell\chi_2(Q_1\|Q_0)}{m}} - \mathbb{Q}^{-1}\left(\alpha + \frac{1}{\sqrt{\ell}}\right)\right) - \frac{1}{\sqrt{\ell}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.223}$$

2.D Proof of Theorem 11

We first prove finite-length converse results.

Theorem 12. *Let w be a positive integer. For any $(M, K, \epsilon, \delta)^d_{\mathrm{avg}}$ and real numbers p, q such that $1/p + 1/q < 1$, we have*

$$\log M \leqslant \frac{1 + w\mathbb{D}(P_1\|P_0)}{1 - \frac{q\epsilon}{1-p(1-\lambda_d(w,\delta))}} + \log\frac{1}{1 - p(1 - \lambda_d(w,\delta))} \tag{2.224}$$

For any $(M, K, \epsilon, \delta)^d_{\max}$ and $\eta > 0$, we have

$$\log M \leqslant \sup_{\mathbf{x}\in\mathcal{X}^n : \boldsymbol{wt}(\mathbf{x})\leqslant w} \mathbb{D}_s^{\epsilon+\eta}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{1-\epsilon}{\eta} + \log\frac{1}{\lambda_d(w,\delta)}. \tag{2.225}$$

Proof. Let $\mathcal{C} = (f, \phi)$ be a code with message set $\mathcal{M}$ and key set $\mathcal{K}$ inducing input distribution $P_{\mathbf{X}}$ over $\mathcal{X}^n$. Let us define $\epsilon_{m,k} \triangleq \mathbb{P}_{Pf(m,k)}(\phi(k, \mathbf{Y}) \neq m)$ for $m \in \mathcal{M}$ and $k \in \mathcal{K}$. If $\mathcal{C}$ is $(|\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta)^d_{\text{avg}}$, then

$$\frac{1}{|\mathcal{M}|\,|\mathcal{K}|} \sum_{m \in \mathcal{M}, k \in \mathcal{K}} \epsilon_{m,k} \leqslant \epsilon, \quad d(W^{\otimes n}_{Z|X} \circ P_{\mathbf{X}}, Q^{\mathbf{0}}) \leqslant \delta. \tag{2.226}$$

Upon defining $\mathcal{S} \triangleq \{(m, k) : \text{wt}(f(m, k)) \leqslant w\}$, we have $|\mathcal{S}|\,/(|\mathcal{M}|\,|\mathcal{K}|) = \mathbb{P}_{P_{\mathbf{X}}}(\mathcal{S}) \geqslant \lambda_d(w, \delta)$. Additionally, by Markov's inequality, there exists $k_0 \in \mathcal{K}$ such that

$$|\mathcal{S}^c \cap \mathcal{M} \times \{k_0\}| \leqslant p\,|S^c|\,/\,|\mathcal{K}| \leqslant p(1 - \lambda_d(w, \delta))\,|\mathcal{M}| \tag{2.227}$$

$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \epsilon_{m,k_0} \leqslant q\epsilon \tag{2.228}$$

We therefore have for $\widetilde{\mathcal{M}} \triangleq \{m \in \mathcal{M} : (m, k_0) \in \mathcal{S}\}$,

$$\left|\widetilde{\mathcal{M}}\right| \geqslant (1 - p(1 - \lambda_d(w, \delta)))\,|\mathcal{M}| \tag{2.229}$$

$$\frac{1}{\left|\widetilde{\mathcal{M}}\right|} \sum_{m \in \widetilde{\mathcal{M}}} \epsilon_{m,k_0} \leqslant \frac{q\epsilon}{1 - p(1 - \lambda_d(w, \delta))}. \tag{2.230}$$

Let $\widetilde{P}_{\mathbf{X}}$ denote the uniform distribution over $f(\widetilde{\mathcal{M}} \times \{k_0\})$. Applying Fano's inequalty, we obtain that

$$\log|\mathcal{M}| + \log(1 - p(1 - \lambda_d(w, \delta))) \leqslant \log\left|\widetilde{\mathcal{M}}\right| \tag{2.231}$$

$$\leqslant \frac{1 + I(\widetilde{P}_{\mathbf{X}}, W^{\otimes n}_{Y|X})}{1 - \frac{q\epsilon}{1 - p(1 - \lambda_d(w,\delta))}} \tag{2.232}$$

$$\leqslant \frac{1 + \mathbb{D}\Big(\widetilde{P}_{\mathbf{X}} \otimes W^{\otimes n}_{Y|X} \| \widetilde{P}_{\mathbf{X}} \otimes P^{\mathbf{0}}\Big)}{1 - \frac{q\epsilon}{1 - p(1 - \lambda_d(w,\delta))}} \tag{2.233}$$

$$\leqslant \frac{1 + \sup_{x \in \text{support}\left(\widetilde{P}_{\mathbf{X}}\right)} \mathbb{D}(P^{\mathbf{x}} \| P^{\mathbf{0}})}{1 - \frac{q\epsilon}{1 - p(1 - \lambda_d(w,\delta))}} \tag{2.234}$$

$$\leqslant \frac{1 + w\mathbb{D}(P_1\|P_0)}{1 - \frac{q\epsilon}{1-p(1-\lambda_d(w,\delta))}}, \tag{2.235}$$

which is the desired bound.

Now suppose that $\mathcal{C}$ is $(|\mathcal{M}|, |\mathcal{K}|, \epsilon, \delta)^d_{\max}$. As before, we define $\mathcal{S}$, for which we have $|\mathcal{S}| \geqslant \lambda_d(w,\delta)\,|\mathcal{M}|\,|\mathcal{K}|$. There exists $k_1 \in \mathcal{K}$ such that $|\mathcal{S} \cap \mathcal{M} \times \{k_1\}| \geqslant |\mathcal{S}|\,/\,|\mathcal{K}| \geqslant \lambda_d(w,\delta)\,|\mathcal{M}|$. Thus,

$$\log|\mathcal{M}| + \log\lambda_d(w,\delta) \leqslant \log|\mathcal{S} \cap \mathcal{M} \times \{k_1\}| \tag{2.236}$$

$$\overset{(a)}{\leqslant} \sup_{\mathbf{x}\in f(\mathcal{S})} -\log\beta_{1-\epsilon}(P^{\mathbf{x}}, P^{\mathbf{0}}) \tag{2.237}$$

$$\overset{(b)}{\leqslant} \sup_{\mathbf{x}\in f(\mathcal{S})} \mathbb{D}_s^{\epsilon+\eta}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{1-\epsilon}{\eta} \tag{2.238}$$

$$\leqslant \sup_{\mathbf{x}\in\mathcal{X}^n:\mathbf{wt}(\mathbf{x})\leqslant w} \mathbb{D}_s^{\epsilon+\eta}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{1-\epsilon}{\eta}, \tag{2.239}$$

where (a) follows form [35, Theorem 31] and (b) follows from [48, Lemma 2.4] □

Consider a sequence of $(M_n, K_n, \epsilon, \delta)^d_{\max}$ codes and let $\gamma_n \triangleq n^{-\frac{1}{2}}$. Using Theorem 12 for $w_n = (\omega - C\gamma_n)n^{\frac{1}{2}}$ and $\eta_n = n^{-\frac{1}{2}}$, we obtain that

$$\log M_n \leqslant \sup_{\mathbf{x}\in\mathcal{X}^n:\mathbf{wt}(\mathbf{x})\leqslant w_n} \mathbb{D}_s^{\epsilon+\eta_n}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + \log\frac{1}{\lambda_d(w_n,\delta)} + \log\frac{1-\epsilon}{\eta_n} \tag{2.240}$$

$$= \sup_{\mathbf{x}\in\mathcal{X}^n:\mathbf{wt}(\mathbf{x})\leqslant w_n} \mathbb{D}_s^{\epsilon+\eta_n}\left(P^{\mathbf{x}}\|P^{\mathbf{0}}\right) + O(\log n) \tag{2.241}$$

$$\overset{(a)}{=} \sup_{w'\in[\![0,w_n]\!]} \mathbb{D}_s^{\epsilon+\eta_n}\left(P_1^{\otimes w'}\|P_0^{\otimes w'}\right) + O(\log n) \tag{2.242}$$

$$\overset{(b)}{\leqslant} \sup_{w'\in[\![1,w_n]\!]} w'\mathbb{D}(P_1\|P_0) - \sqrt{w'V(P_1\|P_0)}\mathbb{Q}^{-1}\left(\epsilon + \eta_n + \frac{6T(P_1\|P_0)}{w'^{\frac{1}{2}}V^{\frac{3}{2}}(P_1\|P_0)}\right) + O(\log n) \tag{2.243}$$

$$= w_n\mathbb{D}(P_1\|P_0) - \sqrt{w_nV(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon) + O(\log n) \tag{2.244}$$

$$= \omega\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} - \sqrt{\omega V(P_1\|P_0)}\mathbb{Q}^{-1}(\epsilon)n^{\frac{1}{4}} + O(\log n), \tag{2.245}$$

where (a) follows from (2.19), and (b) follows from (2.21).

We now consider a sequence of $(M_n, K_n, \epsilon_n, \delta_n)^d_{\text{avg}}$ with $\lim_{n\to\infty} \epsilon_n = 0$ and $\limsup_{n\to\infty} \delta_n = \delta$. Let $\{\gamma_n\}_{n\geqslant 1}$ be such that $\lim_{n\to\infty} \gamma_n = 0$ and set $w_n = (\omega - C\gamma_n)n^{\frac{1}{2}}$. Let $\{p_n\}_{n\geqslant 1}$ and $\{q_n\}_{n\geqslant 1}$ be such that $1/p_n + 1/q_n < 1$ for all n. Applying Theorem 12, we have

$$\log M_n \leqslant \frac{1 + w_n \mathbb{D}(P_1\|P_0)}{1 - \frac{q_n\epsilon_n}{1-p_n(1-\gamma_n)}} + \log \frac{1}{1 - p_n(1-\gamma_n)} \tag{2.246}$$

Setting $q_n = \epsilon_n^{-\frac{1}{2}}$, $p_n = \frac{1}{1-2\epsilon_n^{\frac{1}{2}}}$, and $\gamma_n = 1 - \frac{1-\max(2\epsilon_n^{\frac{1}{4}}, n^{-1})}{p_n}$, we have $1/p_n + 1/q_n = 1 - \epsilon_n^{\frac{1}{2}} < 1$. Furthermore, $\gamma_n \geqslant 1 - \frac{1-2\epsilon_n^{\frac{1}{4}}}{1-2\epsilon_n^{\frac{1}{2}}} > 0$. One can check that γ_n tends to zero as $n \to \infty$, $\log \frac{1}{1-p_n(1-\gamma_n)} \leqslant \log n$, and $\frac{q_n\epsilon_n}{1-p_n(1-\gamma_n)} \leqslant \epsilon_n^{\frac{1}{4}}$. Combining these facts with (2.246), we obtain that

$$\limsup_{n\to\infty} \frac{\log M_n}{\sqrt{n}} \leqslant \omega \mathbb{D}(P_1\|P_0). \tag{2.247}$$

2.E Proof of Lemma 12

Let $A(z) \triangleq \frac{Q_1(z)-Q_0(z)}{Q_0(z)}$ and

$$\mu_x \triangleq \mathbb{E}_{Q_x}(A(Z)), \quad \sigma_x^2 \triangleq \mathbb{E}_{Q_x}\big((A(Z)-\mu_x)^2\big), \quad t_x \triangleq \mathbb{E}_{Q_x}\big(|A(Z)-\mu_x|^3\big) \tag{2.248}$$

Let $\mathcal{T} \triangleq \{\mathbf{z} \in \mathcal{Z}^n : \sum_{i=1}^n A(z_i) \geqslant \tau\}$ for τ determined later. By Theorem 6 in Chapter 1, we have

$$\mathbb{P}_{Q^{\mathbf{o}}}(\mathcal{T}) \leqslant \mathbb{Q}\left(\frac{\tau - n\mu_0}{\sqrt{n}\sigma_0}\right) + \frac{6t_0}{\sqrt{n}\sigma_0^2} \overset{(a)}{=} \mathbb{Q}\left(\frac{\tau}{\sqrt{n}\sigma_0}\right) + \frac{6t_0}{\sqrt{n}\sigma_0^2} \tag{2.249}$$

$$\mathbb{P}_{Q^{\mathbf{x}}}(\mathcal{T}^c) \leqslant 1 - \mathbb{Q}\left(\frac{\tau - n\mu_0 - \mathrm{wt}(\mathbf{x})(\mu_1-\mu_0)}{\sqrt{n\sigma_0^2 + \mathrm{wt}(\mathbf{x})(\sigma_1^2-\sigma_0^2)}}\right) + \frac{6\left(nt_0 + \mathrm{wt}(\mathbf{x})(t_1-t_0)\right)}{(n\sigma_0^2 + \mathrm{wt}(\mathbf{x})(\sigma_1^2-\sigma_0^2))^{\frac{3}{2}}} \tag{2.250}$$

$$\overset{(b)}{=} 1 - \mathbb{Q}\left(\frac{\tau - \mathrm{wt}(\mathbf{x})\mu_1}{\sqrt{n\sigma_0^2 + \mathrm{wt}(\mathbf{x})(\sigma_1^2-\sigma_0^2)}}\right) + \frac{6\left(nt_0 + \mathrm{wt}(\mathbf{x})(t_1-t_0)\right)}{(n\sigma_0^2 + \mathrm{wt}(\mathbf{x})(\sigma_1^2-\sigma_0^2))^{\frac{3}{2}}}, \tag{2.251}$$

where (a) and (b) follows since $\mu_0 = 0$. This yields that

$$\beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}} : \mathbf{x} \in \mathcal{X}^n, \mathrm{wt}(\mathbf{x}) \geqslant w\}) \tag{2.252}$$

$$\leqslant \sup_{w' \geqslant w} 1 - \mathbb{Q}\left(\frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - w'\mu_1}{\sqrt{n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)}}\right) + \frac{6\left(nt_0 + w'(t_1 - t_0)\right)}{\left(n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)\right)^{\frac{3}{2}}} \tag{2.253}$$

$$\leqslant \sup_{w' \geqslant w} 1 - \mathbb{Q}\left(\frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - w'\mu_1}{\sqrt{n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)}}\right) + \frac{6\max(t_0, t_1)}{\sqrt{n}\min(\sigma_0, \sigma_1)}. \tag{2.254}$$

Note further that

$$\frac{\partial}{\partial w'}\left(\frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - w'\mu_1}{\sqrt{n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)}}\right) \tag{2.255}$$

$$= \frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right)(\sigma_1^2 - \sigma_0^2) - \mu_1\left(2n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)\right)}{2\left(n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)\right)^{\frac{3}{2}}}. \tag{2.256}$$

The nominator in (2.256) is negative for large n because the first term is $O(\sqrt{n})$ and the second term is $-O(n)$ and the denominator in (2.256) is positive. Hence, for large n, we have

$$\sup_{w' \geqslant w} 1 - \mathbb{Q}\left(\frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - w'\mu_1}{\sqrt{n\sigma_0^2 + w'(\sigma_1^2 - \sigma_0^2)}}\right) \tag{2.257}$$

$$= 1 - \mathbb{Q}\left(\frac{\sqrt{n}\sigma_0 \mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - w\mu_1}{\sqrt{n\sigma_0^2 + w(\sigma_1^2 - \sigma_0^2)}}\right) \tag{2.258}$$

$$= 1 - \mathbb{Q}\left(\left(\mathbb{Q}^{-1}\left(\alpha - \frac{6t_0}{\sqrt{n}\sigma_0}\right) - \frac{w\mu_1}{\sqrt{n}\sigma_0}\right)\left(1 + \frac{w}{n}\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2}\right)^{-\frac{1}{2}}\right) \tag{2.259}$$

$$= 1 - \mathbb{Q}\left(\left(\mathbb{Q}^{-1}(\alpha) - \frac{w\mu_1}{\sqrt{n}\sigma_0}\right)\right) + O\left(\frac{1}{\sqrt{n}} + \frac{w}{n} + \frac{w^2}{n^{\frac{3}{2}}}\right). \tag{2.260}$$

Finally, straightforward calculations shows that $\mu_1 = \sigma_0^2 = \chi_2(Q_1 \| Q_0)$, which concludes the proof.

CHAPTER 3
FUNDAMENTAL LIMITS OF COVERT COMMUNICATIONS: WIRELESS NON-COHERENT CHANNEL

3.1 Summary

The covert capacity is characterized for a non-coherent fast Rayleigh-fading wireless channel, in which a legitimate user wishes to communicate reliably with a legitimate receiver while escaping detection from a warden. It is shown that the covert capacity is achieved with an amplitude-constrained input distribution that consists of a finite number of mass points including one at zero and numerically tractable bounds are provided. It is also conjectured that distributions with two mass points in fixed locations are optimal. The content of this chapter is based on [49].

3.2 Introduction

For DMCs, the covert-capacity achieving input distribution takes the form of sparse signalling corresponding to those symbols that might arouse suspicion if transmitted, are used a fraction $1/\sqrt{n}$ of the time if n is the block length. Perhaps surprisingly, sparse signalling does *not* achieve the covert-capacity of AWGN channels, as the optimal coding scheme exploits instead Gaussian or BPSK [12] signaling with an average power vanishing as $O\left(1/n\right)$. In other words, encoding information in the *phase* of modulation symbols together with a diffuse power is crucial for optimality. Gaussian signaling has therefore been used to further study covertness over Gaussian and wireless channels, as in [20, 50] to show the benefits of uninformed jammers, in [51] to analyze the role of randomized timing, in [52] to study the effect of randomized power allocation, and in [53] to analyze covert relaying strategies. We note that all aforementioned works exploit random Gaus-

sian codebooks, which simplifies the covertness analysis by reducing the optimal attack to a radiometer. In contrast, we analyze covertness with non-random codebooks using the conceptual approach laid out in [13].

While Gaussian codebooks provides valuable insight into the properties of coding schemes for covert communications over AWGN channels, operating in the vanishing-power regime as suggested by the results might prove challenging. In particular, not only may phase-lock loops fail to properly track the phase of the transmitted signals but symbols with low amplitude may also be severely affected by phase noise, resulting in a significant degradation of the transmission reliability. These effects are also likely to be amplified by the presence of fading in wireless links. The objective of the present paper is to develop insight into this problem by characterizing the covert capacity of non-coherent fast Rayleigh-fading channels (Theorem 13 in Section 3.4), in which the phase is uniformly distributed over $[0; 2\pi[$; although no channel state information is available to the transmitter and receivers, some symbol-level synchronization is assumed.

Our analysis of the covert capacity for non-coherent channels builds upon the ideas initially developed in [54, 55] for amplitude constrained channels and extended to [56] for memoryless non-coherent Rayleigh fading channels under an average power constraint. In particular we show that an optimal covert capacity achieving input distribution is discrete, with one mass point located at zero and subject to an amplitude constraint. While the discrete nature of the distribution may not be a surprise, the fact that the location of the mass points is bounded results from the specific nature of the covertness constraint. We also conjecture that two mass points in *fixed* locations is actually optimal, which is supported by numerical results although we do not have a formal proof. Overall, our results suggest that, in the presence of phase uncertainty, sparse signaling might be an efficient modulation scheme for covert communication.

Our proof technique follows for the most part the high-level approach outlined in [54, 55, 56]; however, the covert communication constraint makes the analysis more intricate

as the optimal capacity-achieving input distribution turns out depend on the block length. In particular, the converse arguments for single-letterization lead to a parameter-dependent constrained optimization problem, in which the parameter should be taken to zero as the blocklength goes to infinity (see the statement of Theorem 13 and (3.18) in Section 3.5). This requires us to analyze the fine dependence of the objective function and the Lagrange multipliers as a function of a parameter using ideas from sensitivity analysis [57].

The rest of this chapter is organized as follows. In Section 3.3, we review the notation in use in this chapter and in Section 3.4, we introduce the precise model for covert communication over non-coherent Rayleigh-fading channels and discuss our characterization of the covert capacity. In Section 3.5, we develop the converse proof of our main result.

3.3 Notation and Conventions

In this chapter, we use continuous probability spaces, which requires its own notation. Let $(\mathcal{S}, \mathcal{F})$ be a measurable space. When $\mathcal{S}$ is a subset of $\mathbb{R}$, we always consider the σ-algebra induced by Borel sets, which converts $\mathcal{S}$ to a measurable space. Let $f : \mathcal{S} \to \mathbb{R}$ be measurable and μ be a measure over $\mathcal{S} \subset \mathbb{R}$. We call f integrable if $\int_{\mathcal{S}} |f| d\mu < \infty$. We then denote the Lebesgue's integral by $\int_{\mathcal{S}} f(x) d\mu$. If $\mathcal{S} =]a, b]$ and μ is the Lebesgue's measure over $\mathcal{S}$, we denote $\int_{\mathcal{S}} f(x) d\mu = \int_a^b f(x) dx = \int_a^b f$. If μ is a probability measure, $X : \mathcal{S} \to \mathbb{R}$ is a random variable, and A is an event, we use $\mathbb{P}_\mu(A)$ and $\mathbb{E}_\mu(X)$ to denote $\mu(A)$ and $\int_{\mathcal{S}} X(s) d\mu$, respectively. When the probability measure μ is discrete, it can be characterized with a PMF $P : \mathcal{S} \to [0, 1]$ satisfying $\mu(A) = \sum_{s \in A} P(s)$. When the probability measure μ is continuous, it can be characterized with a PDF $f : \mathcal{S} \to [0, \infty[$ satisfying $\mu(A) = \int_A f(s) ds$. We do not distinguish between a probability measure and its PMF or PDF (if they exist). The product of two measures μ and μ' is defined in the standard way and is denoted by $\mu \otimes \mu'$. We define the relative entropy between two probability measures μ and μ' as $\mathbb{D}(\mu \| \mu') \triangleq \mathbb{E}_\mu \left(\log \frac{d\mu}{d\mu'} \right)$, where $\frac{d\mu}{d\mu'}$ is the Radon-Nikodym derivative. We also define the χ_2 divergence as $\chi_2(\mu \| \mu') \triangleq \mathbb{E}_{\mu'} \left(\left(\frac{d\mu}{d\mu'} \right)^2 \right) - 1$. We define $\mathbb{I}(X; Y) \triangleq$

$\mathbb{D}(\mu_{XY}\|\mu_X \otimes \mu_Y)$ where μ_{XY}, μ_X and μ_Y denote the probability measures associated to (X, Y), X, and Y, respectively.

Let $\mathcal{X}$ and $\mathcal{Y}$ be two subsets of $\mathbb{R}$. A channel $w_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$ is a mapping $x \mapsto \mu_x$ where μ_x is a probability measure on $\mathcal{Y}$. If μ_x is always continuous, we write $w_{Y|X}(y|x)$ to denote the PDF of μ_x. If μ is a probability measure on $\mathcal{X}$ and $w_{Y|X} : x \mapsto \mu'_x$ is a channel from $\mathcal{X}$ to $\mathcal{Y}$, we define a joint probability measure $w_{Y|X} \otimes \mu$ on $\mathcal{X} \times \mathcal{Y}$ as

$$(\mu \otimes w_{Y|X})(\mathcal{E}) \triangleq \int \mu'_x(\mathcal{E}_x) d\mu, \tag{3.1}$$

where $\mathcal{E}_x \triangleq \{(\widetilde{x}, \widetilde{y}) \in \mathcal{E} : \widetilde{x} = x\}$. We also define the marginal probability measure induced on $\mathcal{Y}$ by $w_{Y|X} \circ \mu$. If X and Y denote the joint random variables associated to the measure $\mu \otimes w_{W|X}$, we allow ourselves to denote their mutual information by $I(\mu, w_{Y|X}) \triangleq \mathbb{I}(X; Y)$.

3.4 System Model

We consider the fast Rayleigh-fading wireless channel illustrated in Fig. 3.4.1, in which at every time instant, the input-output relationships are given by

$$Y = H_m X + N_m \quad \text{and} \quad Z = H_w X + N_w, \tag{3.2}$$

where X is the channel input, Y is the received signal at the legitimate receiver, and Z is the received signal at the warden attempting to detect the transmission. The fading coefficients H_m and H_w are independent complex circular Gaussian random variables with zero-mean and variances θ_m^2 and θ_w^2, respectively. The noises N_m and N_w are also independent zero-mean complex circular random variables with variance σ_m^2 and σ_w^2, respectively. Furthermore, we assume that the channels are stationary and memoryless. The fading coefficients are unknown to all parties, who only have access to their statistical distributions. Since the phase of the fading parameters is uniform, information can only be encoded into

the magnitude of X; additionally, $|Y|^2$ and $|Z|^2$ become sufficient statistics for detection. Hence, as shown in [56], upon re-labeling $|X|^2$ by X and the outputs $|Y|^2$ and $|Z|^2$ by Y and Z, the non-coherent channel is effectively a new memoryless channel with input and output symbols in $[0, \infty[$ and transition probabilities

$$w_{Y|X}(y|x) = \frac{1}{\theta_m^2 x + \sigma_m^2} \exp\left(-\frac{y}{\theta_m^2 x + \sigma_m^2}\right) \tag{3.3}$$

$$w_{Z|X}(z|x) = \frac{1}{\theta_w^2 x + \sigma_w^2} \exp\left(-\frac{z}{\theta_w^2 x + \sigma_w^2}\right). \tag{3.4}$$

By properly normalizing Y and Z, we can assume that $\sigma_w = \sigma_m = 1$, and by normalizing X, we can further assume that $\theta_w = 1$. Thus, we can parameterize the channel by a single parameter[1] θ_m, for which the transition probabilities are

$$p_x(y) \triangleq w_{Y|X}(y|x) = \frac{1}{\theta_m^2 x + 1} \exp\left(-\frac{y}{\theta_m^2 x + 1}\right), \tag{3.5}$$

$$q_x(z) \triangleq w_{Z|X}(z|x) = \frac{1}{x+1} \exp\left(-\frac{z}{x+1}\right). \tag{3.6}$$

Although the input and output sets of the channels are all equal to $[0, \infty[$, we distinguish them with the labels $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ for the input set, the output of main channel, and the output of the warden's channel, respectively.

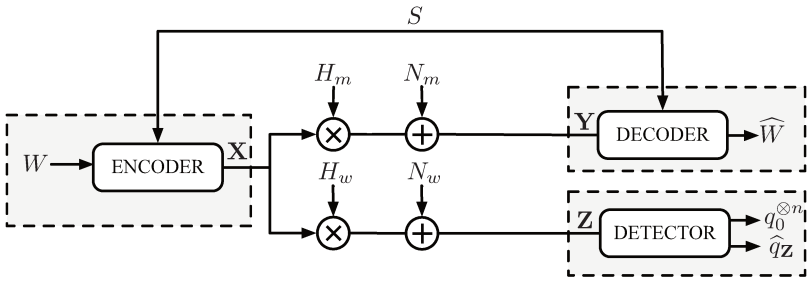

Figure 3.4.1: Covert wireless channel

We next formally describe the covert communication problem in the wireless setting; as depicted in Fig. 3.4.1, the transmitter aims to communicate a message $W \in [\![1, M_n]\!]$

[1]Note that θ_m in (3.4) is different from θ_m in (3.6).

by encoding it into a sequence $\mathbf{X} = (X_1, \cdots, X_n)$ of n symbols using a publicly known coding scheme. Upon observing the corresponding noisy sequence $\mathbf{Y} = (Y_1, \cdots, Y_n)$, the receiver forms an estimate $\widehat{W}$ of W. The encoding and decoding may also use a pre-shared secret key S with an arbitrary distribution over a measurable space.[2] The objective of the warden is to detect the presence of a transmission based on its noisy observation $\mathbf{Z} = (Z_1, \cdots, Z_n)$. The requirements for reliable and covert communication may be formalized as follows. We let $\widehat{q}_{\mathbf{Z}}$ denote the output distribution induced by the coding scheme and $q_0^{\otimes n}$ the product output distribution expected in the absence of communication when the channel input is set to $x = 0$. The performance of an (M_n, n) code transmitting one of M_n message over n channel uses is then measured in terms of the average probability of error $\mathbb{P}(\widehat{W} \neq W)$ and in terms of the relative entropy $\mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n})$.[3][4] Let $\delta > 0$. We say that a covert throughput R is δ-achievable if there exist (M_n, n) codes of increasing block length n such that

$$\log M_n = \omega(\log n), \quad \lim_{n\to\infty} \mathbb{P}(\widehat{W} \neq W) = 0, \tag{3.7}$$

$$\limsup_{n\to\infty} \mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n}) \leqslant \delta, \quad \liminf_{n\to\infty} \frac{\log M_n}{\sqrt{n \mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n})}} \geqslant R. \tag{3.8}$$

The covert capacity, $C_{\text{no-CSI}}(\delta)$, is defined as the supremum of all δ-achievable covert throughputs. Note that we do not specify δ in our terminology of achievable throughput, since it turns out that the normalization of $\log M_n$ in (3.7) removes the dependence on δ.

[2] We show in our achievability proof that a key uniformly distributed over a discrete set with size $O(M_n)$ is sufficient to achieve the covert capacity.

[3] The constraint $\mathbb{D}(\widehat{q}_{\mathbf{Z}} \| q_0^{\otimes n}) \leqslant \delta$ ensures that, regardless of the test performed by the adversary, the sum of the probability of missed detection and false alarm is lower-bounded by $1 - \sqrt{\delta}$. Please refer to [13, Appendix A] for a detailed discussion of the operational meaning of an upper-bound on the relative entropy.

[4] The choice of this specific relative entropy to measure covertness is driven in part by the ease of analysis using channel resolvability techniques. One could of course consider alternative metrics, such as variational distance or a relative entropy with a reversed order of arguments, as discussed in [13, 39]. While the operational meaning of these other metrics remains the same, the analysis and the exact dependence on the constraint δ is metric-specific.

Theorem 13. *Let $\widetilde{\Omega}^{>0}$ be the set of discrete probability measures over $]0,1[$ with a finite number of mass points. $C_{\mathrm{no-CSI}}(\delta)$ is independently of δ equal to*

$$\sup_{\mu\in\widetilde{\Omega}^{>0}} \sqrt{2}\frac{\mathbb{E}_{\mu}(\theta_m^2 X - \log\left(1+\theta_m^2 X\right))}{\sqrt{\mathbb{E}_{\mu\otimes\mu}\left(\frac{X_1X_2}{1-X_1X_2}\right)}}. \tag{3.9}$$

In addition, the following simple bounds hold:

$$\max_{\widetilde{x}\in]0,1]} \widetilde{x}^{-1}\sqrt{2(1-\widetilde{x}^2)}\left(\theta_m^2\widetilde{x} - \log(1+\theta_m^2\widetilde{x})\right) \leqslant C_{\text{no-CSI}}(\delta) \leqslant \sqrt{2}\theta_m^2. \tag{3.10}$$

Theorem 13 provides useful insight into the problem of covert communication over non-coherent channels in several regards. First, a straightforward calculation shows that $\mathbb{D}(p_x\|p_0) = \theta_m^2 x - \log(1+\theta_m^2 x)$ and $\chi_2(w_{Z|X}\circ\mu\|q_0) = \mathbb{E}_{\mu\otimes\mu}\left(\frac{X_1X_2}{1-X_1X_2}\right)$. The expression in (3.9) is therefore a counterpart of [13, Corollary 3] and [12, Eq. (28)]. Second, Theorem 13 shows that we may restrict the signaling schemes for covert communications to finite and amplitude bounded constellations; while the finite nature of the constellation was somewhat expected from the non-coherent nature of the channel, the bound on the amplitude of the points is perhaps more surprising as it was not imposed a priori. We numerically evaluate and plot in Fig. 3.4.2 (3.9) when *the number of mass points in μ is fixed* using a brute-force search. Based on our numerical results, we conjecture that two mass points and OOK signaling is optimal for covert communication.

We omitted the achievability proof, which could be done by pursuing the same approach as in [13, 12]. Please refer to [49] for the complete proof.

3.5 Converse Proof of Theorem 13

Before delving into the detailed proofs, we first provide the sketch of the various steps of the converse proof.

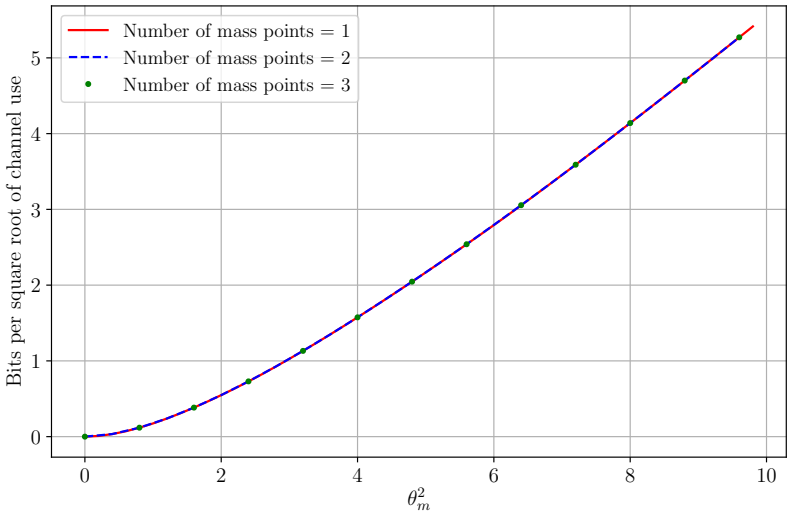

Figure 3.4.2: Numerical evaluation of bounds on covert capacity.

1. We first follow the reasoning of the converse proof of [13] to show that if R is a δ-achievable rate, then there exists a sequence of probability measures $\{\mu_n\}_{n\geqslant}$ over $\mathcal{X}$ such that $\mathbb{D}\big(w_{Z|X} \circ \mu_n \| q_0\big) \leqslant \delta/n$ for n and

$$R \leqslant \liminf_{n\to\infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}\big(w_{Z|X} \circ \mu_n \| q_0\big)}}. \tag{3.11}$$

2. We show that the probability measure μ_n can be further restricted to be discrete with a finite number of mass points and a mass point at zero. This is achieved by investigating the optimization problem

$$\sup_{\mu:\mathbb{D}\left(w_{Z|X}\circ\mu\|q_0\right)\leqslant\nu} I(\mu, w_{Y|X}), \tag{3.12}$$

   and adapting some techniques developed in [56].

3. We prove that we can still upper-bound a covert throughput even if we constraint the amplitude of μ_n as $\max(\text{support}\,(\mu_n)) \leqslant 1 + \zeta$ for any $\zeta > 0$.

4. Let $\{\mu_n\}_n n \geqslant 1$ be a sequence of probability measures such that μ_n has a finite

number of mass and $\max(\text{support}\,(\mu_n)) \leqslant 1+\zeta$. We show that

$$\liminf_{n\to\infty} \frac{I(\mu_n, w_{Y|X})}{\sqrt{\mathbb{D}\big(w_{Z|X}\circ\mu_n \| q_0\big)}} \leqslant \sup_{\mu\in\widetilde{\Omega}^{>0}} \sqrt{2}\frac{\mathbb{E}_\mu(\theta_m^2 X - \log\left(1+\theta_m^2 X\right))}{\sqrt{\mathbb{E}_{\mu\otimes\mu}\Big(\frac{X_1X_2}{1-X_1X_2}\Big)}}. \tag{3.13}$$

3.5.1 Step One: a General Converse for Covert Communication

We consider a sequence of code $\{\mathcal{C}_n\}_{n\geqslant 1}$ where each code $\mathcal{C}_n$ can transmit $\log M_n$ bits with probability of error ϵ_n and relative entropy at most δ_n, and we have $\lim_{n\to\infty}\epsilon_n = 0$ and $\limsup_{n\to\infty}\delta_n \leqslant \delta$. If $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ denotes the input and the output of the channels when $\mathcal{C}_n$ is used and $\widehat{p}_{\mathbf{XYZ}}$ denotes the joint distribution, a standard application of Fano's inequality yields

$$\log M_n \leqslant \frac{\mathbb{I}(\mathbf{X};\mathbf{Y}) + \mathbb{H}_b\left(\epsilon_n\right)}{1-\epsilon_n} \leqslant \frac{\mathbb{I}(\mathbf{X};\mathbf{Y})+1}{1-\epsilon_n}, \tag{3.14}$$

where $\mathbb{H}_b\left(x\right) \triangleq -x\log(x) - (1-x)\log(1-x)$. One can then upper-bound the mutual information $\mathbb{I}(\mathbf{X};\mathbf{Y})$ using standard techniques [58] to obtain

$$\mathbb{I}(\mathbf{X};\mathbf{Y}) \leqslant \sum_{i=1}^{n}\mathbb{I}(X_i;Y_i) \leqslant n\mathbb{I}(\widetilde{X}_n;\widetilde{Y}_n), \tag{3.15}$$

where the random variables $\widetilde{X}_n$ and $\widetilde{Y}_n$ are distributed according to $p_{\widetilde{X}_n}(x) \triangleq \frac{1}{n}\sum_{i=1}^{n}\widehat{p}_{X_i}(x)$ and $p_{\widetilde{X}_n\widetilde{Y}_n}(x,y) \triangleq p_{\widetilde{X}_n}(x)p_x(y)$. Note that $\lim_{n\to\infty} n\mathbb{I}(\widetilde{X}_n;\widetilde{Y}_n) = \infty$ since we assumed that $\log M_n = \omega(\log n)$. Following [59, 12], one can also lower-bound the relative entropy as

$$\delta_n \geqslant \mathbb{D}(\widehat{p}_{\mathbf{Z}}\|q_0^{\otimes n}) \geqslant \sum_{i=1}^{n}\mathbb{D}(\widehat{p}_{Z_i}\|q_0) \geqslant n\mathbb{D}(p_{\widetilde{Z}_n}||q_0), \tag{3.16}$$

where $\widetilde{Z}_n$ is distributed according to $p_{\widetilde{Z}_n}(z) \triangleq \frac{1}{n}\sum_{i=1}^{n} \widehat{p}_{Z_i}(z)$. Consequently,

$$C_{\text{no-CSI}} \leqslant \liminf_{n\to\infty} \frac{\mathbb{I}(\widetilde{X}_n;\widetilde{Y}_n)}{(1-\epsilon_n)\sqrt{\mathbb{D}(p_{\widetilde{Z}_n}\|q_0)}}\left(1+\frac{1}{n\mathbb{I}(\widetilde{X}_n;\widetilde{Y}_n)}\right) = \liminf_{n\to\infty} \frac{\mathbb{I}(\widetilde{X}_n;\widetilde{Y}_n)}{\sqrt{\mathbb{D}(p_{\widetilde{Z}_n}\|q_0)}} \tag{3.17}$$

where the sequence of distributions $\{p_{\widetilde{X}_n\widetilde{Y}_n\widetilde{Z}_n}\}_{n\geqslant 0}$ is subject to the constraint $\mathbb{D}(p_{\widetilde{Z}_n}\|q_0) \leqslant \frac{\delta_n}{n}$. This completes the first step of the converse proof.

3.5.2 Step Two: Discreteness of the Optimal Distribution

We define the optimization problem

$$A(\nu) \triangleq \sup_{\mu\in\Omega:\mathbb{D}\left(w_{Z|X}\circ\mu\|q_0\right)\leqslant\nu} I(\mu, w_{Y|X}), \tag{3.18}$$

where Ω is the set of all probability measures over $\mathcal{X}$ such as μ such that $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) < \infty$. The next lemma shows that there exists a unique maximizer to the above problem.

Lemma 18. *Let $\nu > 0$. There exists a unique probability measure $\mu^*_\nu \in \Omega$ such that $\mathbb{D}\big(w_{Z|X}\circ\mu^*_\nu\|q_0\big) \leqslant \nu$ and $I(\mu^*_\nu, w_{Y|X}) = A(\nu)$.*

Proof. See Appendix 3.E. □

We next characterize the unconstrained form of the optimization in (3.18).

Theorem 14. *Let $\nu > 0$. There exists $\gamma(\nu) \geqslant 0$ such that the following holds.*

1. *We have*

$$A(\nu) = \max\nolimits_{\mu\in\Omega}\left[I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu\right)\right], \tag{3.19}$$

and μ^*_ν is the unique maximizer of the above optimization.

2. *Define*

$$w(x,\mu_1,\nu) \triangleq \int_0^\infty p_x(y)\log\frac{p_x(y)}{(w_{Y|X}\circ\mu_1)(y)}dy - \gamma(\nu)\left(\int_0^\infty q_x(z)\log\frac{(w_{Z|X}\circ\mu_1)(z)}{q_0(z)}dz-\nu\right). \quad (3.20)$$

For all $\mu\in\Omega$, we have

$$A(\nu)\geqslant \mathbb{E}_\mu(w(X,\mu_\nu^*,\nu)). \quad (3.21)$$

3. *Given $\mu_1\in\Omega$, we have for all $\mu\in\Omega$,*

$$A(\nu)\geqslant \mathbb{E}_\mu(w(X,\mu_1,\nu)). \quad (3.22)$$

if and only if

$$w(x,\mu_1,\nu)\leqslant A(\nu)\quad \forall x\in\mathcal{X}, \quad (3.23)$$

$$w(x,\mu_1,\nu)= A(\nu)\quad \forall x\in \text{support}\,(\mu_1). \quad (3.24)$$

4. *We have $\lim_{\nu\to 0^+}\gamma(\nu)=\infty$ and $\lim_{\nu\to 0^+}\gamma(\nu)\nu=0$.*

Proof. See Appendix 3.E. □

Lemma 19. *There exists $\nu_0>0$ such that for all $0<\nu\leqslant\nu_0$, $\text{support}\,(\mu_\nu^*)$ is discrete with a finite number of points in any bounded interval.*

Proof. Fix some $\nu>0$, and define $r(y)\triangleq(w_{Y|X}\circ\mu_\nu^*)(y)$ and $f(z)\triangleq(w_{Z|X}\circ\mu_\nu^*)(z)$. We assume that there exists an interval with an infinite number of points in $\text{support}\,(\mu_\nu^*)$ and obtain a contradiction for ν small enough in four steps.

Step 1: We first use the argument in [56] to show that the KKT condition in (3.24) **holds for all $x\geqslant 0$.** By the Bolzano-Weierstrass theorem, there exists a convergent se-

quence $\{x_i\}_{i\geqslant 1}$ in $\text{support}\,(\mu_\nu^*)$. Moreover, by (3.24), for any $x \in \text{support}\,(\mu_\nu^*)$, we have

$$\phi_\nu(x) \triangleq w(x, \mu_\nu^*, \nu) - A(\nu) \tag{3.25}$$

$$= \int_0^\infty p_x(y) \log \frac{p_x(y)}{r(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{f(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu = 0. \tag{3.26}$$

We now show that $\phi_\nu(x)$ is analytic in x over the domain $\mathcal{D} \triangleq \{x : \mathcal{R}(x) > 0\}$. Note that $\int_0^\infty p_x(y) \log p_x(y) dy = -\log(1+\theta_m x) - 1$ and $\int_0^\infty q_x(z) \log q_0(z) dz = -1 - x$, which are analytic over $\mathcal{D}$. We furthermore have

$$|p_x(y)| = \frac{1}{|1+\theta_m^2 x|} \left| e^{-\frac{y}{1+\theta_m^2 x}} \right| \tag{3.27}$$

$$\overset{(a)}{=} \frac{1}{|1+\theta_m^2 x|} e^{-\frac{y\left(\theta_m^2 \mathcal{R}(x)+1\right)}{|1+\theta_m^2 x|^2}}, \tag{3.28}$$

where (a) follows from $|e^z| = e^{\mathcal{R}(z)}$. This implies that

$$\int_0^\infty |p_x(y) \log r(y)| dy \overset{(a)}{\leqslant} \int_0^\infty |p_x(y)| \left(\theta_m^2 \mathbb{E}_{\mu_\nu^*}(X) + y\right) dy \tag{3.29}$$

$$\overset{(b)}{\leqslant} \int_0^\infty |p_x(y)| \left(\theta_m^2 \left(2\sqrt{\nu} + \nu\right) + y\right) dy \tag{3.30}$$

$$\overset{(c)}{=} \theta_m^2 \left(2\sqrt{\nu} + \nu\right) \frac{|1+\theta_m^2 x|}{\theta_m^2 \mathcal{R}(x) + 1} + \frac{|1+\theta_m^2 x|^3}{(\theta_m^2 \mathcal{R}(x)+1)^2}, \tag{3.31}$$

where (a) follows from (3.139), (b) follows from Lemma 28, and (c) follows by (3.28). Therefore, $\int_0^\infty |p_x(y) \log r(y)| dy$ is uniformly bounded on any compact subset of $\mathcal{D}$, and Theorem 17 yields that $\int_0^\infty |p_x(y) \log r(y)| dy$ is analytic over $\mathcal{D}$. One can similarly argue that $\int_0^\infty q_x(z) \log f(z) dx$ is also analytic over $\mathcal{D}$ and therefore ϕ_ν is analytic. Since $\phi_\nu(x)$ is an analytic function over $\mathcal{D}$, and $\phi_\nu(x) = 0$ over a set with a limit point in $\mathcal{D}$, the identity theorem [60] states that $\phi_\nu(x) = 0$ for all $x \in \mathcal{D}$. Thus, $\phi_\nu(x) = 0$ holds over the entire real line. Using $\int_0^\infty p_x(y) \log p_x(y) dy = -\log(1+\theta_m x) - 1$ and $\int_0^\infty q_x(z) \log q_0(z) dz = -1 - x$,

we can re-write

$$0 = \phi_\nu(x) = -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1+x) - A(\nu) + \gamma(\nu)\nu \\ - \int_0^\infty p_x(y) \log r(y) dy - \gamma(\nu) \int_0^\infty q_x(z) \log f(z) dz. \quad (3.32)$$

To obtain a contradiction, we cannot use the Laplace transform approach of [56] because there are two integrals in (3.32), which is therefore the sum of two Laplace transforms with different arguments. Hence, we continue the proof with another approach.

Step 2: In this step, we shall find the supremum of the support of μ_ν^* in terms of $\gamma(\nu)$. We first consider any non-zero point $\widetilde{x} \in \text{support}\,(\mu_\nu^*)$ and any $\Delta \in]0; \widetilde{x}[$. Since $\widetilde{x} \in \text{support}\,(\mu_\nu^*)$, there exists $\delta > 0$ with $\mu_\nu^*\,(]\widetilde{x} - \Delta, \widetilde{x} + \Delta[) = \delta$. Thus, for any y, by definition of $r(y)$ and the law of total probability, we lower-bound $r(y)$ by

$$r(y) = \mathbb{E}_{\mu_\nu^*}\left(\frac{1}{1+\theta_m^2 X} e^{-\frac{y}{1+\theta_m^2 X}} \right) \quad (3.33)$$

$$\geqslant \mathbb{E}_{\mu_\nu^*}\left(\frac{1}{1+\theta_m^2 X} e^{-\frac{y}{1+\theta_m^2 X}} \middle| X \in]\widetilde{x} - \Delta, \widetilde{x} + \Delta[\right) \mu_\nu^*(]\widetilde{x} - \Delta, \widetilde{x} + \Delta[) \quad (3.34)$$

$$\geqslant \frac{\delta}{1+\theta_m^2(\widetilde{x}+\Delta)} e^{-\frac{y}{1+\theta_m^2(\widetilde{x}-\Delta)}}, \quad (3.35)$$

and similarly, lower-bound $f(z)$ by

$$f(z) \geqslant \frac{\delta}{1+\widetilde{x}+\Delta} e^{-\frac{z}{1+\widetilde{x}-\Delta}}. \quad (3.36)$$

Substituting these bounds in (3.32), we obtain

$$0 \leqslant -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1+x) - A(\nu) + \gamma(\nu)\nu \\ - \int_0^\infty p_x(y) \log \frac{\delta}{1+\theta_m^2(\widetilde{x}+\Delta)} e^{-\frac{y}{1+\theta_m^2(\widetilde{x}-\Delta)}} dy \\ - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{\delta}{1+\widetilde{x}+\Delta} e^{-\frac{y}{1+\widetilde{x}-\Delta}} dz \quad (3.37)$$

$$= -\log(\theta_m^2 x + 1) - 1 - \gamma(\nu)(1+x) - A(\nu) + \gamma(\nu)\nu$$

$$- \log \frac{\delta}{1+\theta_m^2(\widetilde{x}+\Delta)} + \frac{1+\theta_m^2 x}{1+\theta_m^2(\widetilde{x}-\Delta)} - \gamma(\nu)\left(\log \frac{\delta}{1+\widetilde{x}+\Delta} - \frac{1+x}{1+\widetilde{x}-\Delta}\right) \tag{3.38}$$

$$= \kappa - \log(\theta_m^2 x+1) - x\left(\gamma(\nu)\frac{\widetilde{x}-\Delta}{1+\widetilde{x}-\Delta} - \frac{\theta_m^2}{1+\theta_m^2(\widetilde{x}-\Delta)}\right), \tag{3.39}$$

where κ is a constant not depending on x. Since (3.39) holds for all x, by taking the limit $x \to \infty$, we should have

$$\gamma(\nu)\frac{\widetilde{x}-\Delta}{1+\widetilde{x}-\Delta} - \frac{\theta_m^2}{1+\theta_m^2(\widetilde{x}-\Delta)} \leqslant 0. \tag{3.40}$$

Moreover, by letting Δ tend to zero, we obtain

$$\gamma(\nu)\frac{\widetilde{x}}{1+\widetilde{x}} - \frac{\theta_m^2}{1+\theta_m^2\widetilde{x}} \leqslant 0, \tag{3.41}$$

which implies that $x^* \triangleq \sup(\text{support}\,(\mu_\nu^*)) < \infty$. Furthermore, upon finiteness of x^*, we have $r(y) \leqslant e^{-\frac{y}{1+\theta_m^2 x^*}}$, and $f(z) \leqslant e^{-\frac{z}{1+x^*}}$. Replacing these upper-bounds in (3.32), we obtain

$$0 \geqslant -\log(\theta_m^2 x+1) - 1 - \gamma(\nu)(1+x) - A(\nu) + \gamma(\nu)\nu - \int_0^\infty p_x(y)\log e^{-\frac{y}{1+\theta_m^2 x^*}}dy - \gamma(\nu)\int_0^\infty q_x(z)\log e^{-\frac{z}{1+x^*}}dz \tag{3.42}$$

$$= -\log(\theta_m^2 x+1) - 1 - \gamma(\nu)(1+x) - A(\nu) + \gamma(\nu)\nu + \frac{1+\theta_m^2 x}{1+\theta_m^2 x^*} + \gamma(\nu)\frac{1+x}{1+x^*} \tag{3.43}$$

$$= \kappa' - \log(\theta_m^2 x+1) - x\left(\gamma(\nu)\frac{x^*}{1+x^*} - \frac{\theta_m^2}{1+\theta_m^2 x^*}\right), \tag{3.44}$$

where κ' is a constant not depending on x. Since (3.44) holds for all x, we have

$$\gamma(\nu)\frac{x^*}{1+x^*} - \frac{\theta_m^2}{1+\theta_m^2 x^*} \geqslant 0. \tag{3.45}$$

By definition of the support of a distribution, it should be closed, and therefore, $x^* \in \text{support}\,(\mu_\nu^*)$. Since (3.41) holds for all points in the support, we can set $\widetilde{x} = x^*$ and obtain

$$\gamma(\nu)\frac{x^*}{1+x^*} - \frac{\theta_m^2}{1+\theta_m^2 x^*} = 0. \tag{3.46}$$

Step 3: Using the equality for x^* in (3.46), we derive an upper-bound on $A(\nu)$ depending on $\gamma(\nu)$ and ν. By definition of μ_ν^*, it holds that

$$\begin{aligned}
A(\nu) &= I(\mu_\nu^*, w_{Y|X}) && (3.47)\\
&= \mathbb{E}_{w_{Y|X}\otimes\mu_\nu^*}\left(\log\frac{p_X(Y)}{r(Y)}\right) && (3.48)\\
&= \mathbb{E}_{w_{Y|X}\otimes\mu_\nu^*}\left(\log\frac{p_X(Y)p_0(Y)}{r(Y)p_0(Y)}\right) && (3.49)\\
&= \mathbb{E}_{w_{Y|X}\otimes\mu_\nu^*}\left(\log\frac{p_X(Y)}{p_0(Y)}\right) - \mathbb{E}_{w_{Y|X}\circ\mu_\nu^*}\left(\log\frac{r(Y)}{p_0(Y)}\right) && (3.50)\\
&= \mathbb{E}_{w_{Y|X}\otimes\mu_\nu^*}\left(\log\frac{p_X(Y)}{p_0(Y)}\right) - \mathbb{D}(r\|p_0) && (3.51)\\
&\leqslant \mathbb{E}_{w_{Y|X}\otimes\mu_\nu^*}\left(\log\frac{p_X(Y)}{p_0(Y)}\right) && (3.52)\\
&= \mathbb{E}_{\mu_\nu^*}\left(\theta_m^2 X - \log(1+\theta_m^2 X)\right) && (3.53)\\
&\overset{(a)}{\leqslant} \mathbb{E}_{\mu_\nu^*}\left(\frac{1}{2}\theta_m^4 X^2\right) && (3.54)\\
&\leqslant \frac{1}{2}\theta_m^4 x^* \mathbb{E}(X) && (3.55)\\
&\overset{(b)}{\leqslant} \frac{1}{2}\theta_m^4 x^*\left(2\sqrt{\nu}+\nu\right), && (3.56)
\end{aligned}$$

where (a) follows from $\log(1+x) \geqslant x - x^2/2$ for $x \geqslant 0$, and (b) follows from Lemma 28. Therefore, we can use (3.46) to obtain

$$\begin{aligned}
A(\nu) &\leqslant \frac{1}{2}\theta_m^4\left(\frac{\theta_m^4(1+x^*)}{\gamma(\nu)(1+\theta_m^4 x^*)}\right)\left(2\sqrt{\nu}+\nu\right) && (3.57)\\
&\leqslant \frac{2\sqrt{\nu}+\nu}{\gamma(\nu)}\left(\frac{1}{2}\theta_m^4(1+|1-\theta_m^4|)\right). && (3.58)
\end{aligned}$$

Step 4: We complete the proof by obtaining a contradiction. Lemma 33 part 4 implies that there exists $\nu_0 > 0$ and $C > 0$ such that $A(\nu) \geqslant C\sqrt{\nu}$ for all $0 < \nu \leqslant \nu_0$. By Theorem 14 part 4, we can choose ν_0 small such that $\gamma(\nu) > \frac{3}{C}\left(\frac{1}{2}\theta_m^4(1+|1-\theta_m^4|)\right)$ in addition to $A(\nu) \geqslant C\sqrt{\nu}$ for all $0 < \nu \leqslant \nu_0$. Since by decreasing ν_0, the statement would be weaker, we can always assume that $\nu_0 < 1$. Thus,

$$
\begin{aligned}
C\sqrt{\nu} &\leqslant A(\nu) && (3.59)\\
&\leqslant \frac{2\sqrt{\nu}+\nu}{\gamma(\nu)}\left(\frac{1}{2}\theta_m^4(1+|1-\theta_m^4|)\right) && (3.60)\\
&< \frac{2\sqrt{\nu}+\nu}{\frac{3}{C}\left(\frac{1}{2}\theta_m^4(1+|1-\theta_m^4|)\right)}\left(\frac{1}{2}\theta_m^4(1+|1-\theta_m^4|)\right) && (3.61)\\
&\leqslant C\sqrt{\nu}. && (3.62)
\end{aligned}
$$

□

Lemma 20. *There exists $\nu_0 > 0$ such that for any $\nu_0 > \nu > 0$, the support of μ_ν^* has a finite number of points.*

Proof. We proceed by contradiction. Assume that the support of μ_ν^* has infinitely many points $\{x_i\}_{i=1}^{\infty}$ in increasing order with probabilities $\{\alpha_i\}_{i=1}^{\infty}$. Since we proved that in any bounded interval, we can only have a finite number of points, $\lim_{i\to\infty} x_i = \infty$. Note that for any $j \geqslant 1$, we have

$$
\begin{aligned}
(w_{Y|X} \circ \mu_\nu^*)(y) &= \sum_{i=1}^{\infty} \alpha_i p_{x_i}(y) && (3.63)\\
&\geqslant \alpha_j p_{x_j}(y), && (3.64)
\end{aligned}
$$

and

$$
\left(w_{Z|X} \circ \mu_\nu^*\right)(z) \geqslant \alpha_j q_{x_j}(z). \tag{3.65}
$$

Therefore, for all $j \geqslant 1$, we can upper-bound $\phi_\nu(x)$ defined in (3.25) as

$$\phi_\nu(x) = \int_0^\infty p_x(y) \log \frac{p_x(y)}{(w_{Y|X} \circ \mu_\nu^*)(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{(w_{Z|X} \circ \mu_\nu^*)(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu \tag{3.66}$$

$$\leqslant \int_0^\infty p_x(y) \log \frac{p_x(y)}{\alpha_j p_{x_j}(y)} dy - \gamma(\nu) \int_0^\infty q_x(z) \log \frac{\alpha_j q_{x_j}(z)}{q_0(z)} dz - A(\nu) + \gamma(\nu)\nu \tag{3.67}$$

$$= \log(\theta_m^2 x + 1) - 1 - \log \frac{\alpha_j}{1 + \theta_m^2 x_j} + \frac{1 + \theta_m^2 x}{1 + \theta_m^2 x_j} - \gamma(\nu) \left(1 + x - \log \frac{\alpha_j}{1 + x_j} + \frac{1 + x}{1 + x_j} \right) - A(\nu) + \gamma(\nu)\nu \tag{3.68}$$

$$= \kappa + \log(\theta_m^2 x + 1) + \left(-\gamma(\nu) + \frac{\gamma(\nu)}{1 + x_j} + \frac{\theta_m^2}{1 + \theta_m^2 x_j} \right) x, \tag{3.69}$$

where κ is a constant not depending on x. Furthermore, the KKT condition in (3.24) implies that (3.69) is non-negative for all x_i, and since x_i can be large enough, we should have

$$-\gamma(\nu) + \frac{\gamma(\nu)}{1 + x_j} + \frac{\theta_m^2}{1 + \theta_m^2 x_j} \geqslant 0. \tag{3.70}$$

Because x_j can be large enough, we have $-\gamma(\nu) \geqslant 0$. This cannot be true for small ν since $\lim_{\nu \to 0^+} \gamma(\nu) = \infty$ by Theorem 14. □

Lemma 21. *There exists $\nu_0 > 0$ such that for all $\nu_0 > \nu > 0$, μ_ν^* has a mass point at 0.*

The proof of Lemma 21 will require the following technical result which is a modification of [56, Lemma 1].

Lemma 22. *Let $f(z)$ be a PDF with mean m and $g(z)$ be a strictly monotonically increasing function, then $\int (z - m) f(z) g(z) dz > 0$.*

Proof. $(z - m)(g(z) - g(m))$ is always positive as either the product of two negative terms if $z < m$ or two positive terms if $z > m$. Thus, $(z - m)g(z) > (z - m)g(m)$ and

$\int (z-m)g(z)f(z)dz > \int (z-m)g(m)f(z)dz = 0.$ $\square$

Proof of Lemma 21. Let ν_0 be as in Lemma 20 so that μ_ν^* has finite number of mass points for all $0 < \nu \leqslant \nu_0$. For the sake of a contradiction, assume that μ_ν^* is a discrete probability measure over $\mathcal{X}$ with k mass points $0 < x_1 < \cdots < x_k$ with corresponding probabilities $\alpha_1, \cdots, \alpha_k$. In [56], it is proved that reducing x_1 increases the mutual information $I(\mu, w_{Y|X})$. Therefore, to complete the proof, it is enough to show that $\frac{\partial \mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right)}{\partial x_1} > 0$. Defining $f(x_1, z) \triangleq (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)}$, we have

$$\frac{\partial \mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right)}{\partial x_1} = \frac{\partial}{\partial x_1} \int_{\mathcal{Z}} f(x_1, z) dz. \tag{3.71}$$

By Lemma 30, $\int_{\mathcal{Z}} |f(x_1, z)| dz < \infty$, and we have

$$\frac{\partial f}{\partial x_1}(x_1, z) = \frac{\alpha_1}{(1+x_1)^2} q_{x_1}(z) \left(z - \mathbb{E}_{q_{x_1}}(Z)\right) \left(\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} + 1\right), \tag{3.72}$$

which satisfies that

$$\left| \frac{\partial f}{\partial x_1}(x_1, z) \right| \leqslant e^{-\frac{z}{1+x_1}} (z + x_1 + 1) \left(2z + \mathbb{E}_\mu(X) + 1\right). \tag{3.73}$$

The right hand side of (3.73), is bounded with an integrable function of z independent of x_1, if x_1 is bounded. Hence, Theorem 16 implies that

$$\frac{\partial \mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right)}{\partial x_1} = \alpha_1 \frac{1}{(1+x_1)^2} \int_0^\infty (z - \mathbb{E}_{q_{x_1}}(Z)) q_{x_1}(z) \left(\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} + 1\right) dz. \tag{3.74}$$

Note that

$$\log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} = \log \frac{\sum_{i=1}^{k} \alpha_i \frac{1}{x_i+1} e^{-\frac{z}{x_i+1}}}{e^{-z}} \tag{3.75}$$

$$= \log \sum_{i=1}^{k} \alpha_i \frac{1}{x_i+1} e^{z\frac{x_i}{x_i+1}}. \tag{3.76}$$

Since $1 > \frac{1}{x_1+1} > \cdots > \frac{1}{x_k+1}$, $\log \frac{(w_{Z|X}\circ\mu)(z)}{q_0(z)} + 1$ is strictly monotonically increasing in z. Using Lemma 22, $\frac{\partial \mathbb{D}\left(w_{Z|X}\circ\mu \| q_0\right)}{\partial x_1} > 0$, and hence, by decreasing x_1, the constraint $\mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right) \leqslant \nu$ still holds and $I(\mu, w_{Y|X})$ is increased. This contradicts with the definition of μ_ν^*, and therefore, there exists a mass point at zero. □

3.5.3 Step Three: an Amplitude Constraint

For a probability measure μ on $\mathcal{X}$ and $a > 0$, we define $\mathbb{C}_a[\mu]$ as a new probability measure $\widetilde{\mu}$ on $\mathcal{X}$ such that

$$\widetilde{\mu}]-\infty, x[) = \begin{cases} \mu]-\infty, x[) & x < a, \\ 1 & x \geqslant a. \end{cases} \tag{3.77}$$

Intuitively, $\widetilde{\mu}$ is obtained by moving all probability of $]a, \infty[$ in μ to a mass point at a.

Theorem 15. *Let $\{\nu_n\}_{n\geqslant 1}$ be $o(1)$. For all $a > 1$, if n is large enough, we have $\mathbb{C}_a[\mu_{\nu_n}^*] \in \Omega_a(\nu_n)$ and*

$$\liminf_{n\to\infty} \frac{I(\mu_{\nu_n}^*, w_{Y|X})}{\sqrt{\mathbb{D}\left(w_{Z|X}\circ\mu_{\nu_n}^* \| q_0\right)}} \leqslant \liminf_{n\to\infty} \frac{I(\mathbb{C}_a[\mu_{\nu_n}^*], w_{Y|X})}{\sqrt{\mathbb{D}\left(w_{Z|X}\circ\mathbb{C}_a[\mu_{\nu_n}^*] \| q_0\right)}}. \tag{3.78}$$

To prove this result, we need the following lemmas.

Lemma 23. *If μ is a discrete probability measure on $\mathcal{X}$ with finite number of mass points*

$x_1 < \cdots < x_k$ *and corresponding probabilities* $\alpha_1, \cdots, \alpha_k$*, then*

$$\mathbb{D}\big(w_{Z|X} \circ \mathbb{C}_a[\mu] \| q_0\big) \leqslant \mathbb{D}\big(W_{Z|X} \circ \mu \| q_0\big), \tag{3.79}$$

$$I(\mathbb{C}_a[\mu], w_{Y|X}) \geqslant I(\mu, w_{Y|X}) - \theta_m^2 \text{max}(\text{support}\,(\mu))\mu(]a, \infty[). \tag{3.80}$$

Proof. Similar to (3.74), for all $i \in [\![1, k]\!]$, we have

$$\frac{\partial}{\partial x_i}\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) = \alpha_i \frac{1}{(1+x_i)^2} \int_0^\infty (z - \mathbb{E}_{q_{x_i}}(Z)) q_{x_i}(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \geqslant 0. \tag{3.81}$$

Hence, by moving all mass points located in $]a, \infty[$ to a to obtain $\mathbb{C}_a[\mu]$, we decrease the relative entropy. Applying the same argument to the channel $w_{Y|X}$, we have $\mathbb{D}\big(w_{Y|X} \circ \mathbb{C}_a[\mu] \| p_0\big) \leqslant \mathbb{D}\big(w_{Y|X} \circ \mu \| p_0\big)$. Additionally, we have

$$I(\mu, w_{Y|X}) = \sum_{i=1}^k \alpha_i \mathbb{D}(p_{x_i} \| p_0) - \mathbb{D}\big(w_{Y|X} \circ \mu \| p_0\big) \tag{3.82}$$

$$= \sum_{i=1}^k \alpha_i \left(\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)\right) - \mathbb{D}\big(w_{Y|X} \circ \mu \| p_0\big), \tag{3.83}$$

which implies that

$$I(\mu, w_{Y|X}) - I(\mathbb{C}_a[\mu], w_{Y|X}) \tag{3.84}$$

$$= \left(\sum_{i: x_i > a} \alpha_i \left(\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)\right) - \mu(]a, \infty[) \left(\theta_m^2 a - \log(1 + \theta_m^2 a)\right)\right) +$$

$$\big(-\mathbb{D}\big(w_{Y|X} \circ \mu \| p_0\big) + \mathbb{D}\big(w_{Y|X} \circ \mathbb{C}_a[\mu] \| p_0\big)\big) \tag{3.85}$$

$$\leqslant \sum_{i: x_i > a} \alpha_i \left(\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)\right) - \mu(]a, \infty[) \left(\theta_m^2 a - \log(1 + \theta_m^2 a)\right) \tag{3.86}$$

$$\leqslant \sum_{i: x_i > a} \alpha_i \left(\theta_m^2 x_i - \log(1 + \theta_m^2 x_i)\right) \tag{3.87}$$

$$\leqslant \theta_m^2 \text{max}(\text{support}\,(\mu))\mu(]a, \infty[). \tag{3.88}$$

□

Lemma 24. *For all $a > 0$, there exist $\nu_0 > 0$, $\widetilde{x} \in \mathcal{X}$, and $\xi > 0$ such that for all $0 < \nu \leqslant \nu_0$, if $\max(\text{support}\,(\mu_v^*)) \geqslant \widetilde{x}$, then $\mu_v^*(]a, \infty[) \leqslant 2^{-\xi \max(\text{support}(\mu_v^*))}$.*

Proof. Fix $\nu > 0$ small enough and suppose that $\mu \triangleq \mu_\nu^*$ has mass points $x_1 < \cdots < x_k$ with corresponding probabilities $\alpha_1, \cdots, \alpha_k$. Let $r(y) \triangleq (w_{Y|X} \circ \mu)(z)$ and $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$. Substituting the lower-bounds

$$r(y) \geqslant \frac{\mu(]a, \infty[)}{1 + \theta_m^2 x_k} e^{-\frac{y}{1+\theta_m^2 a}}, \text{ and } f(z) \geqslant \frac{\mu(]a, \infty[)}{1 + x_k} e^{-\frac{z}{1+a}}, \tag{3.89}$$

in the KKT condition (3.24) for the point $x = x_k$, we obtain

$$\begin{aligned} 0 \leqslant &- \log(\theta_m^2 x_k + 1) - 1 - \gamma(\nu)(1 + x_k) - A(\nu) + \gamma(\nu)\nu - \\ &\log \frac{\mu(]a, \infty[)}{1 + \theta_m^2 x_k} + \frac{1 + \theta_m^2 x_k}{1 + \theta_m^2 a} + \gamma(\nu) \left(- \log \frac{\mu(]a, \infty[)}{1 + x_k} + \frac{1 + x_k}{1 + a} \right). \end{aligned} \tag{3.90}$$

Since $\lim_{\nu \to 0^+} \gamma(\nu)\nu = 0$, for small ν, $-1 - A(\nu) + \gamma(\nu)\nu \leqslant 0$, and therefore, (3.90) implies that

$$0 \leqslant -\gamma(\nu)(1 + x_k)\frac{a}{1 + a} + \gamma(\nu) \log(1 + x_k) + \frac{1 + \theta_m^2 x_k}{1 + \theta_m^2 a} - (1 + \gamma(\nu)) \log(\mu(]a, \infty[)). \tag{3.91}$$

Furthermore, if x_k is large enough, we have $\log(1 + x_k) \leqslant \frac{(1+x_k)a}{4(1+a)}$, and if ν is small enough and x_k is large enough, by Theorem 14 part 4, we have $\frac{1+\theta_m^2 x_k}{1+\theta_m^2 a} \leqslant \gamma(\nu)(1 + x_k)\frac{a}{4(1+a)}$. Hence, there exist $\nu_0 > 0$ and $\widetilde{x} > 0$ such that if $\nu \leqslant \nu_0$ and $x_k \geqslant \widetilde{x}$, we have

$$0 \leqslant -\frac{1}{2}\gamma(\nu)(1 + x_k)\frac{a}{1 + a} - (1 + \gamma(\nu)) \log(\mu(]a, \infty[)), \tag{3.92}$$

which yields that

$$\mu(]a,\infty[) \leqslant \exp\left(-\frac{1}{2}\frac{\gamma(\nu)}{1+\gamma(\nu)}(1+x_k)\frac{a}{1+a}\right). \tag{3.93}$$

Since $\lim_{\nu\to 0^+}\gamma(\nu)=\infty$, there exists $\nu_0>0$ such that $\inf_{\nu\in]0,\nu_0]}\frac{\gamma(\nu)}{1+\gamma(\nu)}\geqslant\frac{1}{2}$. Hence, for $\xi\triangleq\frac{a}{4(1+a)}$ and all $0<\nu<\nu_0$, we have $\mu(]a,\infty[)\leqslant 2^{-\xi x_k}$. □

We are now ready to establish the upper bound in (3.10) of Theorem 15.

Proof of Theorem 15. Let $x_n^*\triangleq\max(\text{support}\left(\mu_{\nu_n}^*\right))$. By Lemma 20, if n is large enough $\mu_{\nu_n}^*$ is a discrete probability measure with finite number of mass points, and so is $\mathbb{C}_a[\mu_{\nu_n}^*]$. By Lemma 23, we have $\mathbb{D}\big(w_{Z|X}\circ\mathbb{C}_a[\mu_{\nu_n}^*]\|q_0\big)\leqslant\mathbb{D}\big(\mu_{\nu_n}^*\|q_0\big)=\nu_n$, and

$$\frac{I(\mathbb{C}_a[\mu_{\nu_n}^*],w_{Y|X})}{\sqrt{\mathbb{D}\big(w_{Z|X}\circ\mathbb{C}_a[\mu_{\nu_n}^*]\|q_0\big)}}\geqslant\frac{I(\mu_{\nu_n}^*,w_{Y|X})-\theta_m^2x_n^*\mu_{\nu_n}^*(]a,\infty[)}{\sqrt{\mathbb{D}\big(w_{Z|X}\circ\mu_{\nu_n}^*\|q_0\big)}}. \tag{3.94}$$

Therefore, it is enough to show that

$$x_n^*\mu_{\nu_n}^*(]a,\infty[)=o\left(\sqrt{\mathbb{D}\big(w_{Z|X}\circ\mu_{\nu_n}^*\|q_0\big)}\right)=o\left(\sqrt{\nu_n}\right). \tag{3.95}$$

To do so, we consider ν_0, $\widetilde{x}$, and ξ from Lemma 24. For n large enough such that $\frac{2}{\xi}\log\frac{1}{\nu_n}>\widetilde{x}$, if $x_n^*\geqslant\frac{2}{\xi}\log\frac{1}{\nu_n}$, then

$$x_n^*\mu_{\nu_n}^*(]a,\infty[)\leqslant x_n^*2^{-\xi x_n^*}, \tag{3.96}$$

which is less than $2^{-\frac{1}{2}\xi x_n^*}$ for large enough n. Thus, $x_n^*\geqslant\frac{2}{\xi}\log\frac{1}{\nu_n}$ implies that $x_n^*\mu_{\nu_n}^*(]a,\infty[)\leqslant\frac{1}{\nu_n}$. For the other case when $x_n^*<\frac{2}{\xi}\log\frac{1}{\nu_n}$, let $\widetilde{\mu}$ be a probability distribution on $\mathcal{X}$ with two mass points at 0 and a with probabilities $1-\mu_{\nu_n}^*(]a,\infty[)$ and $\mu_{\nu_n}^*(]a,\infty[)$, respectively. Then, we have

$$\nu_n=\mathbb{D}\big(w_{Z|X}\circ\mu_{\nu_n}^*\|q_0\big)\overset{(a)}{\geqslant}\mathbb{D}\big(w_{Z|X}\circ\widetilde{\mu}\|q_0\big)\overset{(b)}{\geqslant}K\left(\mu_{\nu_n}^*(]a,\infty[)\right)^{\frac{a+1}{a}}, \tag{3.97}$$

where (a) follows from the same argument as in the proof of Lemma 23, and (b) follows from Lemma 32 for a constant K depending on a. Therefore, we have

$$x_n^* \mu_{\nu_n}^*(]a,\infty[) \leqslant \frac{2}{\xi} \log \frac{1}{\nu_n} \left(\frac{\nu_n}{K}\right)^{\frac{a}{a+1}}. \tag{3.98}$$

Since both $\frac{1}{\nu_n}$ and $\frac{2}{\xi} \log \frac{1}{\nu_n} \left(\frac{\nu_n}{K}\right)^{\frac{a}{a+1}}$ are $o(\sqrt{\nu_n})$, we have (3.95). □

3.5.4 Step Four: Obtaining the Bound in Theorem 13

We first prove a lemma that relates the constraint on the relative entropy to χ_2 divergence. Let $\widetilde{\Omega}^{\geqslant 0}$ be the set of discrete probability measures over $[0,1[$ with finite number of mass points.

Lemma 25. *Let $\epsilon > 0$ be small enough and $\{\nu_n\}_{n\geqslant}$ be a sequence of real numbers such that $\lim_{n\to\infty} \nu_n = 0$ and $2\sqrt{\nu_n} + \nu_n \leqslant 0.5$ for all n. There exists a sequence of probability measures $\{\lambda_n\}_{n\geqslant 1}$ such that $\lambda_n \in \widetilde{\Omega}^{\geqslant 0}$ and*

$$\limsup_{n\to\infty} \frac{A(\nu_n)}{\sqrt{\nu_n}} \leqslant \limsup_{n\to\infty} \frac{I(\lambda_n, w_{Y|X})}{\sqrt{\frac{1}{2}\chi_2(w_{Z|X} \circ \lambda_n \| q_0)}} + \epsilon. \tag{3.99}$$

Proof. Let $\xi > 0$ and $\zeta \triangleq \frac{6\xi}{1-6\xi}$. Define

$$\mu_n \triangleq \mu_{\nu_n}^* \tag{3.100}$$

$$\mu_n' \triangleq C_{1+\zeta}[\mu_n] \tag{3.101}$$

$$\mu_n'' \triangleq C_{a_n}[\mu_n'], \tag{3.102}$$

where $a_n' \triangleq \inf_{a:\mu_n(]a,\infty[) \leqslant \nu_n^{\frac{1}{2}+\xi}} a$ and $a_n \triangleq \min(1-\zeta, a_n')$. Let $\{\nu_n\}_{n\geqslant}$ be a sequence of real numbers such that $\lim_{n\to\infty} \nu_n = 0$ and $2\sqrt{\nu_n} + \nu_n \leqslant 0.5$ for all n. By construction, we have $\mu_n''([a_n', \infty[) \geqslant \nu_n^{\frac{1}{2}+\xi}$ and $\mu_n''(]a_n', \infty[) \leqslant \nu_n^{\frac{1}{2}+\xi}$. We next use the following lemma to upper-bound $\chi_2(w_{Z|X} \circ \mu_n'' \| q_0)$.

Lemma 26. *Let $\mu \in \widetilde{\Omega}^{\geqslant 0}$ such that $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) \leqslant \nu$ and $\max\left(\text{support}\left(\mu\right)\right) \leqslant a < 1$. If $2\sqrt{\nu} + \nu < 1/2$ and for some $M > 0$, we have $(w_{Z|X} \circ \mu)(M)/q_0(M) \geqslant e$, then*

$$\frac{1}{2}\chi_2(w_{Z|X} \circ \mu \| q_0) \leqslant \mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) + \frac{1}{2}(\mathbb{E}_{\mu}(X))^3 \int_0^M e^{z\left(-1+\frac{3a}{1+a}\right)} dz + \frac{1}{2}(\mathbb{E}_{\mu}(X))^2 \int_M^{\infty} e^{z\left(-1+\frac{2a}{1+a}\right)} dz + 2(\mathbb{E}_{\mu}(X))^3. \quad (3.103)$$

Proof. See Appendix 3.D. □

We first establish a lower-bound on $(w_{Z|X} \circ \mu_n''(z))/q_0(z)$ to use Lemma 26. Since $a_n \leqslant a_n'$, we have

$$\mu_n''([a_n, \infty[) \geqslant \mu_n''([a_n', \infty[) \geqslant \nu_n^{\frac{1}{2}+\xi}, \quad (3.104)$$

which yields that

$$\frac{(w_{Z|X} \circ \mu_n'')(z)}{q_0(z)} \geqslant \nu_n^{\frac{1}{2}+\xi} \frac{e^{\frac{a_n}{1+a_n} z}}{1+a_n}. \quad (3.105)$$

Choosing $M_n = \frac{1+a_n}{a_n}\left(2 + \left(\frac{1}{2}+\zeta\right)\log\frac{1}{\nu_n}\right)$, we have $(w_{Z|X} \circ \mu_n'')(M_n)/q_0(M_n) \geqslant e$. Therefore, Lemma 26 implies that

$$\begin{aligned}
&\frac{1}{2}\chi_2(w_{Z|X} \circ \mu_n'' \| q_0) && (3.106)\\
&\leqslant \mathbb{D}\big(w_{Z|X} \circ \mu_n'' \| q_0\big) + \frac{1}{2}(\mathbb{E}_{\mu_n''}(X))^3 \int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz \\
&\qquad + \frac{1}{2}(\mathbb{E}_{\mu_n''}(X))^2 \int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz + 2(\mathbb{E}_{\mu_n''}(X))^3 && (3.107)\\
&\overset{(a)}{\leqslant} \nu_n + \frac{1}{2}(\mathbb{E}_{\mu_n''}(X))^3 \int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz \\
&\qquad + \frac{1}{2}(\mathbb{E}_{\mu_n''}(X))^2 \int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz + 2(\mathbb{E}_{\mu_n''}(X))^3 && (3.108)\\
&\overset{(b)}{\leqslant} \nu_n \left(1 + \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz + \frac{9}{2}\int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz + 27\nu_n^{\frac{1}{2}}\right), && (3.109)
\end{aligned}$$

where (a) follows since by Lemma 23

$$\mathbb{D}\big(w_{Z|X} \circ \mu_n'' \| q_0\big) \leqslant \mathbb{D}\big(w_{Z|X} \circ \mu_n' \| q_0\big) \leqslant \mathbb{D}\big(w_{Z|X} \circ \mu_n \| q_0\big) \leqslant \nu_n, \tag{3.110}$$

and (b) follows since by Lemma 28, we have $\mathbb{E}_{\mu_n''}(X) \leqslant 2\sqrt{\nu_n} + \nu_n \leqslant 3\sqrt{\nu_n}$. We now show that

$$\lim_{n\to\infty} \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz = \lim_{n\to\infty} \frac{9}{2} \int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz = \lim_{n\to\infty} 27 M_0 \nu_n^{\frac{1}{2}} = 0. \tag{3.111}$$

For the first limit, we consider two cases.

If $a_n \leqslant 1/4$, then $-1 + \frac{3a_n}{1+a_n} \leqslant -2/5$ and

$$\frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz \leqslant \frac{27}{2}\nu_n^{\frac{1}{2}} \int_0^{\infty} e^{-\frac{2z}{5}} dz \tag{3.112}$$

$$= \frac{135}{4}\nu_n^{\frac{1}{2}}. \tag{3.113}$$

If $a_n \geqslant 1/4$, then

$$\int_0^{M_n} e^{z\left(-1+\frac{3a_n}{1+a_n}\right)} dz \leqslant M_n \max\left(1, e^{M_n\left(-1+\frac{3a_n}{1+a_n}\right)}\right) \tag{3.114}$$

$$= \frac{1+a_n}{a_n}\left(2 + \left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right) \times \max\left(1, e^{\frac{1+a_n}{a_n}\left(2+\left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right)\left(-1+\frac{3a_n}{1+a_n}\right)}\right) \tag{3.115}$$

$$\leqslant 5\left(2 + \left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right) \max\left(1, e^{\frac{2a_n-1}{a_n}\left(2+\left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right)}\right). \tag{3.116}$$

Note that

$$5\left(2 + \left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right) = O\left(\log\frac{1}{\nu_n}\right) \tag{3.117}$$

and

$$\max\left(1, e^{\frac{2a_n-1}{a_n}\left(2+\left(\frac{1}{2}+\xi\right)\log\frac{1}{\nu_n}\right)}\right) = O\left(1+\nu_n^{-\frac{2a_n-1}{a_n}\left(\frac{1}{2}+\xi\right)}\right) \tag{3.118}$$

For $a_n \leqslant 1 - \frac{6\xi}{1-6\xi}$, we have $\frac{2a_n-1}{a_n}\left(\frac{1}{2}+\xi\right) \leqslant \frac{1}{2} - \xi$. For the second limit in (3.111), note that

$$\int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz \leqslant \int_{M_n}^{\infty} e^{-z\frac{\zeta}{2-\zeta}} dz. \tag{3.119}$$

Because $\int_0^{\infty} e^{-z\frac{\xi}{2-\xi}} < \infty$ and $M_n = \frac{1+a_n}{a_n}\left(2+\left(\frac{1}{2}+\zeta\right)\log\frac{1}{\nu_n}\right)$ goes to infinity as n goes to infinity, $\lim_{n\to\infty}\int_{M_n}^{\infty} e^{z\left(-1+\frac{2a_n}{1+a_n}\right)} dz = 0$. The third limit in (3.111) follows since $\lim_{n\to\infty}\nu_n = 0$. We thus obtain (3.111), which together with (3.109) results in

$$\limsup_{n\to\infty} \frac{\frac{1}{2}\chi_2(w_{Z|X}\circ\mu_n''\|q_0)}{\nu_n} \leqslant 1. \tag{3.120}$$

We now consider $I(\mu_n'', w_{Y|X})$ and show that it is close to $I(\mu_n, w_{Y|X}) = A(\nu_n)$.

If $a_n = 1-\zeta$, then by a modification of Lemma 23

$$\begin{aligned} I(\mu_n'', w_{Z|X}) &\geqslant I(\mu_n', w_{Z|X}) - 2\zeta\mu_n'(]1-\zeta,\infty[) && (3.121)\\ &= I(\mu_n', w_{Z|X}) - 2\zeta\mu_n(]1-\zeta,\infty[) && (3.122)\\ &\geqslant I(\mu_n', w_{Z|X}) - 2\zeta\frac{\mathbb{E}_{\mu_n}(X)}{1-\zeta} && (3.123)\\ &\geqslant I(\mu_n', w_{Z|X}) - 6\zeta\frac{\sqrt{\nu_n}}{1-\zeta}. && (3.124) \end{aligned}$$

If $a_n = a_n'$, by Lemma 23

$$\begin{aligned} I(\mu_n'', w_{Z|X}) &\geqslant I(\mu_n', w_{Z|X}) - 2(1+\zeta)\mu_n'(]a_n',\infty[) && (3.125)\\ &= I(\mu_n', w_{Z|X}) - 2(1+\zeta)\nu_n^{\frac{1}{2}+\zeta}. && (3.126) \end{aligned}$$

Therefore,

$$I(\mu_n'', w_{Z|X}) \geqslant I(\mu_n', w_{Z|X}) - \max\left(6\zeta\frac{\sqrt{\nu_n}}{1-\zeta}, 2(1+\zeta)\nu_n^{\frac{1}{2}+\zeta}\right) \tag{3.127}$$

$$\overset{(a)}{\geqslant} I(\mu_n, w_{Z|X}) - \max\left(6\zeta\frac{\sqrt{\nu_n}}{1-\zeta}, 2(1+\zeta)\nu_n^{\frac{1}{2}+\zeta}\right) - o(\nu_n^{\frac{1}{2}}), \tag{3.128}$$

where (a) follows from the argument of Theorem 15. Taking $\lambda_n = \mu_n'' \in \widetilde{\Omega}^{\geqslant 0}$, by (3.128) and (3.120), we have (3.99) for $\epsilon = \frac{6\zeta}{1-\zeta}$. □

Let $\mu \in \widetilde{\Omega}^{\geqslant 0}$. We claim that

$$\frac{I(\mu, w_{Y|X})}{\sqrt{\chi_2(w_{Z|X} \circ \mu \| q_0)}} \leqslant \sup_{\widetilde{\mu} \in \widetilde{\Omega}^{>0}} \frac{\mathbb{E}_{\widetilde{\mu}}(\theta_m^2 X - \log(1+\theta_m^2 X))}{\sqrt{\mathbb{E}_{\widetilde{\mu}\otimes\widetilde{\mu}}\left(\frac{X_1X_2}{1-X_1X_2}\right)}}. \tag{3.129}$$

Let us define $\widetilde{\mu}$ as

$$\widetilde{\mu}(A) \triangleq \frac{\mu(A \cap]0,1[)}{\mu(]0,1[)}. \tag{3.130}$$

In other words, $\widetilde{\mu}$ is the probability measure μ conditioned to the event $]0,1[$. We have

$$I(\mu, w_{Y|X}) \leqslant \mathbb{E}_{\mu}\big(\theta_m^2 X - \log\big(1+\theta_m^2 X\big)\big) \overset{(a)}{=} \mu(]0,1[)\mathbb{E}_{\widetilde{\mu}}\big(\theta_m^2 X - \log\big(1+\theta_m^2 X\big)\big), \tag{3.131}$$

where (a) follows since $\theta_m^2 x - \log(1+\theta_m^2 x) = 0$ for $x = 0$. Moreover,

$$\chi_2(w_{Z|X} \circ \mu \| q_0) = \mathbb{E}_{\mu\circ\mu}\left(\frac{X_1X_2}{1-X_1X_2}\right) = \mu(]0,1[)^2\mathbb{E}_{\widetilde{\mu}\circ\widetilde{\mu}}\left(\frac{X_1X_2}{1-X_1X_2}\right) \tag{3.132}$$

Therefore,

$$\frac{I(\mu, w_{Y|X})}{\sqrt{\frac{1}{2}\chi_2(w_{Z|X} \circ \mu \| q_0)}} \leqslant \sqrt{2}\frac{\mathbb{E}_{\widetilde{\mu}}(\theta_m^2 X - \log(1+\theta_m^2 X))}{\sqrt{\mathbb{E}_{\widetilde{\mu}\otimes\widetilde{\mu}}\left(\frac{X_1X_2}{1-X_1X_2}\right)}}. \tag{3.133}$$

Furthermore, with the help of Lemma 33, Eq. (3.265), we have that

$$\limsup_{\nu \to 0^+} \frac{A(\nu)}{\sqrt{\nu}} \leqslant \sqrt{2}\theta_m^2. \tag{3.134}$$

Therefore, we obtain the upper-bound in (3.10).

3.6 Conclusion

For covert communications over non-coherent wireless channels, we showed that discrete constellations with an amplitude constraint are optimal. This differs from the results for coherent Gaussian channels in which using the phase is required to achieve the covert capacity. Supported by numerical results, we also conjectured that the optimal number of points is two and that their positions are fixed.

APPENDIX

3.A Leibniz Integral Rule

For a reader's convenience, we recall Leibniz integral rule here as it is used extensively throughout the paper.

Theorem 16. *Let $\mathcal{O}$ be an open subset of $\mathbb{R}$ and Ω be a measure space. Suppose $f : \mathcal{O} \times \Omega \to \mathbb{R}$ satisfies the following conditions*

1. *$f(x, w)$ is a Lebesgue-integrable function of ω for each $x \in \mathcal{O}$*
2. *For almost all $\omega \in \Omega$, the derivative $\frac{\partial f}{\partial x}$ exists for all $x \in \mathcal{O}$*
3. *There is an integrable function $\theta : \Omega \to \mathbb{R}$ such that $\left|\frac{\partial f}{\partial x}(x, \omega)\right| \leqslant \theta(\omega)$ for all $x \in \mathcal{O}$ and almost every $\omega \in \Omega$.*

Then, for all $x \in \mathcal{O}$, we have

$$\frac{d}{dx}\int f(x,\omega)d\omega = \int \frac{\partial f}{\partial x}(x,\omega)d\omega \tag{3.135}$$

3.B An Analyticity Criterion

Theorem 17. *Let $g : \mathcal{D} \times \mathbb{R} \to \mathbb{C}$ be a function such that $\mathcal{D}$ is a simple connected subset of $\mathbb{C}$, $g(\cdot, y)$ is analytic for all $y \in \mathbb{R}$, and $\sup_{z\in\mathcal{C}} \int_{\mathbb{R}} |g(z, y)|dy < \infty$ for all compact $\mathcal{C} \subset \mathcal{D}$. The function $f : z \mapsto \int_{\mathbb{R}} g(z, y)dy$ is analytic over the domain $\mathcal{D}$.*

Proof. The proof is a straightforward application of Fubini's theorem and Morera's theo-

rem. Fixing any closed piecewise C^1 curve γ in $\mathcal{D}$, we have

$$\int_\gamma f(z)dz = \int_\gamma \int_{\mathbb{R}} g(z,y)dydz \tag{3.136}$$

$$\overset{(a)}{=} \int_{\mathbb{R}} \int_\gamma g(z,y)dzdy \tag{3.137}$$

$$\overset{(b)}{=} 0, \tag{3.138}$$

where (a) follows from Fubini's theorem and our assumption on g, and (b) follows since $g(\cdot, z)$ is analytic and from Cauchy's integral theorem. Therefore, f satisfies the condition of Morera's theorem and is analytic. □

3.C Auxiliary Results

We gather here essential technical tools to prove the achievability and converse results. To begin with, we bound the PDF of the output distributions of the channels $w_{Y|X}$ and $w_{Z|X}$ for an arbitrary input distribution μ.

Proposition 2. *For any probability measure μ on $\mathcal{X}$ with $\mathbb{E}_\mu(X) < \infty$ and all $y \in \mathcal{Y}, z \in \mathcal{Z}$, we have*

$$-\theta_m^2 \mathbb{E}_\mu(X) - y \leqslant \log((w_{Y|X} \circ \mu)(y)) \leqslant 0, \tag{3.139}$$

$$-\mathbb{E}_\mu(X) - z \leqslant \log((w_{Z|X} \circ \mu)(z)) \leqslant 0, \tag{3.140}$$

$$\mathbb{E}_{w_{Y|X}\circ\mu}(Y) = 1 + \theta_m^2 \mathbb{E}_\mu(X), \tag{3.141}$$

$$\mathbb{E}_{w_{Z|X}\circ\mu}(Z) = 1 + \mathbb{E}_\mu(X). \tag{3.142}$$

Proof. We only prove (3.139) and (3.141), from which (3.140) and (3.142) follow by setting $\theta_m = 1$. To obtain (3.139), observe that for any $x \in \mathcal{X}$, we have $p_x(y) \triangleq$

$\frac{1}{1+\theta_m^2 x} e^{-\frac{y}{1+\theta_m^2 x}} \leqslant 1$, and

$$\log\big((w_{Y|X} \circ \mu)(y)\big) = \log\left(\mathbb{E}_\mu(p_X(y))\right) \tag{3.143}$$

$$\overset{(a)}{\geqslant} \mathbb{E}_\mu(\log\left(p_X(y)\right)) \tag{3.144}$$

$$= \mathbb{E}_\mu\left(-\log\left(1+\theta_m^2 X\right) - \frac{y}{1+\theta_m^2 X}\right) \tag{3.145}$$

$$\overset{(b)}{\geqslant} \mathbb{E}_\mu\left(-\theta_m^2 X - \frac{y}{1+\theta_m^2 X}\right) \tag{3.146}$$

$$\overset{(c)}{\geqslant} \mathbb{E}_\mu\big(-\theta_m^2 X - y\big) \tag{3.147}$$

$$= -\theta_m^2 \mathbb{E}_\mu(X) - y, \tag{3.148}$$

where (a) follows from Jensen's inequality, (b) follows from $\log(1+x) \leqslant x$ for $x > -1$, and (c) follows from $\mathbb{P}_\mu(X \geqslant 0) = 1$. To obtain (3.141), note that

$$\mathbb{E}_{w_{Y|X}\circ\mu}(Y) = \int_0^\infty y(w_{Y|X} \circ \mu)(y)dy \tag{3.149}$$

$$= \int_0^\infty y\left(\int_{\mathcal{X}} p_x(y)d\mu\right)dy \tag{3.150}$$

$$\overset{(a)}{=} \int_{\mathcal{X}}\left(\int_0^\infty y p_x(y)dy\right)d\mu \tag{3.151}$$

$$= \int_{\mathcal{X}}(1+\theta_m^2 x)d\mu \tag{3.152}$$

$$= 1 + \theta_m^2 \mathbb{E}_\mu(X), \tag{3.153}$$

where (a) follows from Fubini's theorem and the fact that for all x, y, $y p_x(y) \geqslant 0$. □

Lemma 27. *Let μ be a probability measure over $\mathcal{X}$. If $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big)$ exists and is finite, then $\mathbb{E}_\mu(X) < \infty$.*

Proof. We proceed by contradiction. Consider a positive real number γ_1 and let $2\epsilon \triangleq \mu([\gamma_1, \infty)$. We have $\epsilon > 0$, because otherwise $\mathbb{E}_\mu(X) \leqslant \gamma_1 < \infty$. By the continuity of a

probability, we have

$$\lim_{\gamma\to\infty}\mu([\gamma_1,\gamma])=2\epsilon. \tag{3.154}$$

Therefore, there exists $\gamma_2\geqslant\gamma_1$ such that $\mu([\gamma_1,\gamma_2])\geqslant\epsilon$. We then have

$$(w_{Z|X}\circ\mu)(z)\geqslant\frac{\epsilon e^{-\frac{z}{1+\gamma_1}}}{1+\gamma_2}. \tag{3.155}$$

This implies that $(w_{Z|X}\circ\mu)(z)\geqslant q_0(z)=e^{-z}$ for all $z\geqslant z_0\triangleq\frac{1+\gamma_1}{\gamma_1}\log\frac{1+\gamma_2}{\epsilon}>0$. Since $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big)<\infty$, we have

$$\begin{aligned}
\infty&>\int_{z_0}^{\infty}(w_{Z|X}\circ\mu)(z)\log\frac{(w_{Z|X}\circ\mu)(z)}{q_0(z)}dz && (3.156)\\
&\geqslant\int_{z_0}^{\infty}(w_{Z|X}\circ\mu)(z)\log\frac{\frac{\epsilon e^{-\frac{z}{1+\gamma_1}}}{1+\gamma_2}}{e^{-z}}dz && (3.157)\\
&\geqslant\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\int_{z_0}^{\infty}(w_{Z|X}\circ\mu)(z)zdz && (3.158)\\
&=\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\int_{z_0}^{\infty}\int_{\mathcal{X}}\frac{e^{-\frac{z}{1+x}}}{1+x}d\mu zdz && (3.159)\\
&=\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\int_{\mathcal{X}}\int_{z_0}^{\infty}\frac{e^{-\frac{z}{1+x}}}{1+x}zdzd\mu && (3.160)\\
&=\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\int_{\mathcal{X}}(1+x)\left(1+\frac{z_0}{1+x}\right)e^{-\frac{z_0}{1+x}}d\mu && (3.161)\\
&\geqslant\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\int_{\mathcal{X}}(1+x)e^{-z_0}d\mu && (3.162)\\
&\geqslant\log\frac{\epsilon}{1+\gamma_2}+\frac{\gamma_1}{1+\gamma_1}\left(\mathbb{E}_\mu(X)+1\right)e^{-z_0}, && (3.163)
\end{aligned}$$

which implies that $\mathbb{E}_\mu(X)<\infty$. □

The next result shows that an upper-bound on $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big)$ leads to an upper-bound on $\mathbb{E}_\mu(X)$.

Lemma 28. *For any $\nu>0$ and for any probability measure μ on $\mathcal{X}$, $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big)\leqslant\nu$ implies that $\mathbb{E}_\mu(X)\leqslant 2\sqrt{\nu}+\nu$.*

Proof. For any $x \in \mathbb{R}^+$, we first consider the relative entropy $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_x\big)$ and show that it exists. By (3.140) in Proposition 2 applied to a distribution with a single mass point at x, $|\log q_x(z)| \leqslant x+z$. We thus have $\int_0^\infty (w_{Z|X} \circ \mu)(z)\, |\log q_x(z)|\, dz \leqslant x + \mathbb{E}_{w_{Z|X}\circ\mu}(Z) = x + 1 + \mathbb{E}_\mu(X)$, which is finite by Lemma 27. Consequently, $\int_0^\infty (w_{Z|X} \circ \mu)(z) \log q_x(z) dz$ is finite, and therefore by [61, Lemma 8.3.1], the relative entropy $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_x\big)$ exists and is finite. Accordingly, we have

$$0 \geqslant -\int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_x(z)} dz \tag{3.164}$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) \left(-\log((w_{Z|X} \circ \mu)(z)) - \log(1+x) - \frac{z}{1+x} \right) dz. \tag{3.165}$$

Furthermore, by our assumption that $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) \leqslant \nu$, we have

$$\nu \geqslant \int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \tag{3.166}$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) \left(\log((w_{Z|X} \circ \mu)(z)) + z \right) dz. \tag{3.167}$$

Adding the inequalities in (3.165) and (3.167), we obtain

$$\nu \geqslant \int_0^\infty (w_{Z|X} \circ \mu)(z) \left(-\log(1+x) + \frac{xz}{1+x} \right) dz \tag{3.168}$$

$$= -\log(1+x) + \frac{x}{1+x} \mathbb{E}_{w_{Z|X}\circ\mu}(Z) \tag{3.169}$$

$$\overset{(a)}{=} -\log(1+x) + \frac{x}{1+x} \left(\mathbb{E}_\mu(X) + 1 \right), \tag{3.170}$$

where (a) follows from (3.142). Hence, we have

$$\mathbb{E}_\mu(X) \leqslant (\nu + \log(1+x)) \frac{1+x}{x} - 1 \tag{3.171}$$

$$\leqslant (\nu + x) \frac{1+x}{x} - 1. \tag{3.172}$$

Choosing $x = \sqrt{\nu}$, we obtain the desired upper-bound. □

Lemma 29. *For any probability measure μ on $\mathcal{X}$ with $\mathbb{E}_\mu(X) < \infty$, $I(\mu, w_{Y|X})$ is well-defined and finite, and*

$$I(\mu, w_{Y|X}) = -\int_0^\infty (w_{Y|X} \circ \mu)(y) \log((w_{Y|X} \circ \mu)(y)) dy - \mathbb{E}_\mu\big(\log\left(1 + \theta_m^2 X\right)\big) - 1. \tag{3.173}$$

Proof. To check that $I(\mu, w_{Y|X})$ is well-defined and finite, it is enough to show that

$$\int \left| \log \frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right| d(w_{Y|X} \otimes \mu) < \infty, \tag{3.174}$$

which holds since

$$\begin{aligned}
&\int \left| \log \frac{p_x(y)}{(w_{Y|X} \circ \mu)(y)} \right| d(w_{Y|X} \otimes \mu) && (3.175)\\
&\qquad \leqslant \int \big(|\log p_x(y)| + |\log((w_{Y|X} \circ \mu)(y))| \big)\, d(w_{Y|X} \otimes \mu) && (3.176)\\
&\qquad \overset{(a)}{\leqslant} \int \big(\theta_m^2 (x + \mathbb{E}_\mu(X)) + 2y\big)\, d(w_{Y|X} \otimes \mu) && (3.177)\\
&\qquad = 2\theta_m^2 \mathbb{E}_\mu(X) + 2\mathbb{E}_{w_{Y|X} \circ \mu}(Y) && (3.178)\\
&\qquad \overset{(b)}{=} 4\theta_m^2 \mathbb{E}_\mu(X) + 2 < \infty, && (3.179)
\end{aligned}$$

where (a) follows from (3.139), and (b) follows from (3.141). Note next that

$$\begin{aligned}
I(\mu, w_{Y|X}) &= \mathbb{E}_{w_{Y|X} \otimes \mu}\left(\log \frac{p_X(Y)}{(w_{Y|X} \circ \mu)(Y)} \right) && (3.180)\\
&= \mathbb{E}_{w_{Y|X} \otimes \mu}\left(-\log(1 + \theta_m^2 X) - \frac{Y}{1 + \theta_m^2 X} - \log((w_{Y|X} \circ \mu)(Y)) \right). && (3.181)
\end{aligned}$$

Moreover, $\mathbb{E}(\log(1 + \theta_m^2 X)) \leqslant \theta_m^2 \mathbb{E}(X) < \infty$ and $\mathbb{E}\left(\frac{Y}{1+\theta_m^2 X}\right) \leqslant \mathbb{E}(Y) < \infty$, and there-

fore, we can use the linearity of expectation to write

$$\mathbb{E}_{w_{Y|X}\otimes\mu}\bigg(-\log(1+\theta_m^2X)-\frac{Y}{1+\theta_m^2X}-\log\big((w_{Y|X}\circ\mu)(Y)\big)\bigg) \tag{3.182}$$

$$=-\mathbb{E}\big(\log(1+\theta_m^2X)\big)-\mathbb{E}\bigg(\frac{Y}{1+\theta_m^2X}\bigg)-\mathbb{E}\big(\log\big((w_{Y|X}\circ\mu)(Y)\big)\big) \tag{3.183}$$

$$=-\mathbb{E}\big(\log(1+\theta_m^2X)\big)-\mathbb{E}\bigg(\mathbb{E}\bigg(\frac{Y}{1+\theta_m^2X}\bigg|X\bigg)\bigg)-\mathbb{E}\big(\log\big((w_{Y|X}\circ\mu)(Y)\big)\big) \tag{3.184}$$

$$=-\mathbb{E}\big(\log(1+\theta_m^2X)\big)-\mathbb{E}\bigg(\frac{1+\theta_m^2X}{1+\theta_m^2X}\bigg)-\mathbb{E}\big(\log\big((w_{Y|X}\circ\mu)(Y)\big)\big) \tag{3.185}$$

$$=-\mathbb{E}\big(\log(1+\theta_m^2X)\big)-1-\mathbb{E}\big(\log\big((w_{Y|X}\circ\mu)(Y)\big)\big), \tag{3.186}$$

which completes the proof of (3.173). □

Lemma 30. *Suppose that* $\mathbb{D}\big(w_{Z|X}\circ\mu_1\|q_0\big)$ *and* $\mathbb{D}\big(w_{Z|X}\circ\mu_2\|q_0\big)$ *exist and are finite for two probability measures* μ_1 *and* μ_2 *on* $\mathcal{X}$*. Then, the cross entropy* $\int_0^\infty(w_{Z|X}\circ\mu_1)(z)\log(w_{Z|X}\circ\mu_2(z))dz$ *exists and is finite.*

Proof. We shall show that $\int_0^\infty(w_{Z|X}\circ\mu_1)(z)|\log((w_{Z|X}\circ\mu_2)(z))|dz<\infty$. By Lemma 28, we know that $\mathbb{E}_{\mu_1}(X)$ and $\mathbb{E}_{\mu_2}(X)$ are finite. Therefore, we have

$$\int_0^\infty(w_{Z|X}\circ\mu_1)(z)|\log(w_{Z|X}\circ\mu_2(z))|dz\overset{(a)}{\leqslant}\int_0^\infty(w_{Z|X}\circ\mu_1)(z)\left(\mathbb{E}_{\mu_2}(X)+z\right)dz \tag{3.187}$$

$$=\mathbb{E}_{\mu_2}(X)+\mathbb{E}_{w_{Z|X}\circ\mu_1}(Z) \tag{3.188}$$

$$\overset{(b)}{=}\mathbb{E}_{\mu_2}(X)+1+\mathbb{E}_{\mu_1}(X)<\infty \tag{3.189}$$

where (a) follows from (3.140), and (b) follows from (3.142). □

Lemma 31. *Let* μ *be a probability measure over* $\mathcal{X}$ *such that* $\sup(\text{support}\,(\mu))<\infty$*. We then have*

$$I(\mu,w_{Y|X})=\mathbb{E}_\mu\big(\theta_m^2X-\log(1+\theta_m^2X)\big)-\mathbb{D}\big(w_{Y|X}\circ\mu\|p_0\big). \tag{3.190}$$

Furthermore, if we have $\sup(\text{support}\,(\mu)) < 1$*, then*

$$\chi_2(w_{Z|X} \circ \mu \| q_0) = \mathbb{E}_{\mu\circ\mu}\bigg(\frac{X_1X_2}{1-X_1X_2}\bigg). \tag{3.191}$$

Proof. We have

$$\begin{aligned}
I(\mu, w_{Y|X}) &= \int \log \frac{p_X(Y)}{(w_{Y|X}\circ\mu)(Y)} d(w_{Y|X}\otimes\mu) && (3.192)\\
&= \int \log \frac{p_X(Y)}{p_0(Y)} d(w_{Y|X}\otimes\mu) + \int \log \frac{p_0(Y)}{(w_{Y|X}\circ\mu)(Y)} d(w_{Y|X}\otimes\mu) && (3.193)\\
&= \mathbb{E}_\mu(\mathbb{D}(p_X\|p_0)) - \mathbb{D}\big(w_{Y|X}\circ\mu\|p_0\big) && (3.194)\\
&\overset{(a)}{=} \mathbb{E}_\mu\big(\theta_m^2 X - \log(1+\theta_m^2 X)\big) - \mathbb{D}\big(w_{Y|X}\circ\mu\|p_0\big), && (3.195)
\end{aligned}$$

where (a) follows from the straightforward calculation of the relative entropy between two exponential distribution. Additionally, we have

$$\begin{aligned}
\chi_2(w_{Z|X}\circ\mu\|q_0) &= \int_0^\infty \frac{(w_{Z|X}\circ\mu)(z)^2}{q_0(z)} dz - 1 && (3.196)\\
&= \int_0^\infty \mathbb{E}_{\mu\otimes\mu}\bigg(\frac{1}{(1+X_1)(1+X_2)} e^{z\left(1-\frac{1}{1+X_1}-\frac{1}{1+X_2}\right)}\bigg) dz - 1 && (3.197)\\
&\overset{(a)}{=} \mathbb{E}_{\mu\otimes\mu}\bigg(\int_0^\infty \frac{1}{(1+X_1)(1+X_2)} e^{z\left(1-\frac{1}{1+X_1}-\frac{1}{1+X_2}\right)} dz\bigg) - 1 && (3.198)\\
&= \mathbb{E}_{\mu\otimes\mu}\bigg(\frac{1}{1-X_1X_2}\bigg) - 1 && (3.199)\\
&= \mathbb{E}_{\mu\otimes\mu}\bigg(\frac{X_1X_2}{1-X_1X_2}\bigg), && (3.200)
\end{aligned}$$

where (a) follows from Fubini theorem and $\frac{1}{(1+X_1)(1+X_2)} e^{z\left(1-\frac{1}{1+X_1}-\frac{1}{1+X_2}\right)} \geqslant 0$ almost surely. □

Lemma 32. *If $a > 1$ and $\beta > 0$ is small enough, then*

$$\mathbb{D}(\beta q_a + (1-\beta)q_0 \| q_0) = \\ \beta^{1+\frac{1}{a}} (1+a)^{-1-\frac{1}{a}} \left(1+\frac{1}{a}\right) \left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2} + a^2 \Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)\right) + O(\beta^2), \tag{3.201}$$

where $\Gamma(x) \triangleq \int_0^\infty y^{x-1} e^{-y} dy$.

If $a < 1$ and $\beta > 0$ is small enough, then

$$\mathbb{D}(\beta q_a + (1-\beta)q_0 \| q_0) = \frac{a^2}{2(1-a^2)} \beta^2 + o(\beta^2). \tag{3.202}$$

Proof. We only consider the case where $a > 1$ and the other case follows from similar approach. By definition, we have

$$\mathbb{D}(\beta q_a + (1-\beta)q_0 \| q_0) \tag{3.203}$$
$$= \int_0^\infty (\beta q_a(z) + (1-\beta)q_0(z)) \log\left(\frac{\beta q_a(z) + (1-\beta)q_0(z)}{q_0(z)}\right) dz \tag{3.204}$$
$$= \int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1-\beta)e^{-z}\right) \log\left(1-\beta+\frac{\beta}{1+a} e^{\frac{az}{1+a}}\right) dz \tag{3.205}$$
$$= \log(1-\beta) + \int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1-\beta)e^{-z}\right) \log\left(1+\frac{\beta}{(1-\beta)(1+a)} e^{\frac{az}{1+a}}\right) dz. \tag{3.206}$$

By substitution $u \triangleq e^{\frac{az}{1+a}}$ in the above integral, we obtain

$$\int_0^\infty \left(\beta \frac{e^{-\frac{z}{1+a}}}{1+a} + (1-\beta)e^{-z}\right) \log\left(1+\frac{\beta}{(1-\beta)(1+a)} e^{\frac{az}{1+a}}\right) dz = \\ \left(1+\frac{1}{a}\right) \int_1^\infty \left((1-\beta)u^{-2-\frac{1}{a}} + \frac{\beta}{1+a} u^{-1-\frac{1}{a}}\right) \log\left(1+\frac{\beta}{(1-\beta)(1+a)} u\right) du \tag{3.207}$$

Note next that for all real numbers λ_1, λ_2, a primitive function of $u^{\lambda_1} \log(1+\lambda_2 u)$ up to a

constant is

$$\frac{u^{\lambda_1+1}\left({}_2F_1(1,\lambda_1+1;\lambda_1+2;-\lambda_2 u)+(\lambda_1+1)\log(\lambda_2 u+1)-1\right)}{(\lambda_1+1)^2}, \tag{3.208}$$

where ${}_2F_1(a,b;c;x)$ is the hypergeometric function. Additionally, for $\lambda_1 < -1$, the limit of this primitive function at $u=\infty$ is

$$\frac{\lambda_2^{-\lambda_1-1}\Gamma(2+\lambda_1)\Gamma(-\lambda_1)}{(\lambda_1+1)^2}. \tag{3.209}$$

Therefore, if we define $\lambda \triangleq \frac{\beta}{(1-\beta)(1+a)}$, by linearity of integral, we have

$$\int_1^\infty \left((1-\beta)u^{-2-\frac{1}{a}}+\frac{\beta u^{-\frac{1}{a}}}{1+a}\right)\log\left(1+\frac{\beta}{(1-\beta)(1+a)}u\right)du \tag{3.210}$$

$$= (1-\beta)\left(\frac{\lambda^{1+\frac{1}{a}}\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}-\frac{{}_2F_1\left(1,-1-\frac{1}{a};-\frac{1}{a};-\lambda\right)-(1+\frac{1}{a})\log(\lambda+1)-1}{\left(1+\frac{1}{a}\right)^2}\right)$$
$$+\frac{\beta}{1+a}\left(\frac{\lambda^{\frac{1}{a}}\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}-\frac{{}_2F_1\left(1,-\frac{1}{a};1-\frac{1}{a};-\lambda\right)-(\frac{1}{a})\log(\lambda+1)-1}{\left(\frac{1}{a}\right)^2}\right) \tag{3.211}$$

$$\overset{(a)}{=} (1-\beta)\left(\frac{\lambda^{1+\frac{1}{a}}\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}-\frac{\left(1+\frac{\lambda(1+\frac{1}{a})}{-\frac{1}{a}}\right)-(1+\frac{1}{a})\lambda-1+O(\beta^2)}{\left(1+\frac{1}{a}\right)^2}\right) \tag{3.212}$$

$$+\frac{\beta}{1+a}\left(\frac{\lambda^{\frac{1}{a}}\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}-\frac{\left(1+\frac{\lambda^{\frac{1}{a}}}{1-\frac{1}{a}}\right)-(\frac{1}{a})\lambda+O(\beta^2)-1}{\left(\frac{1}{a}\right)^2}\right) \tag{3.213}$$

$$= (1-\beta)\left(\frac{\lambda^{-1-\frac{1}{a}}\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}-\frac{\frac{\lambda(1+\frac{1}{a})}{-\frac{1}{a}}-(1+\frac{1}{a})\lambda+O(\beta^2)}{\left(1+\frac{1}{a}\right)^2}\right) \tag{3.214}$$

$$+\lambda(1-\beta)\left(\frac{\lambda^{-\frac{1}{a}}\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}-\frac{\frac{\lambda^{\frac{1}{a}}}{1-\frac{1}{a}}-(\frac{1}{a})\lambda+O(\beta^2)}{\left(\frac{1}{a}\right)^2}\right), \tag{3.215}$$

where (a) follows since for x going to zero ${}_2F_1(a,b;c;x) = 1 + abx/c + O(x^2)$ and $\log(1+x) = x+O(x^2)$ by Taylor's expansion. By rearranging the terms in above expression and disregarding the higher order terms, we obtain

$$\begin{aligned}
&\lambda^{1+\frac{1}{a}}(1-\beta)\left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}+\frac{\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}\right)+\lambda\frac{1+a}{1+\frac{1}{a}}(1-\beta)+O(\beta^2)\\
&=\beta^{1+\frac{1}{a}}\left(\frac{1}{(1-\beta)(1+a)}\right)^{1+\frac{1}{a}}(1-\beta)\left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}+\frac{\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}\right)\\
&\qquad+\frac{\beta}{1+\frac{1}{a}}+O(\beta^2). \quad (3.216)
\end{aligned}$$

Combining (3.206), (3.207), and (3.216), we have

$$\begin{aligned}
\mathbb{D}(\beta q_a+(1-\beta)q_0\|q_0) &= \beta^{1+\frac{1}{a}}\left(\frac{1}{(1-\beta)(1+a)}\right)^{1+\frac{1}{a}}\\
&\times(1-\beta)(1+\frac{1}{a})\left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(2+\frac{1}{a}\right)}{\left(1+\frac{1}{a}\right)^2}+\frac{\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(1+\frac{1}{a}\right)}{\left(\frac{1}{a}\right)^2}\right)+O(\beta^2). \quad (3.217)
\end{aligned}$$

□

3.D Proof of Lemma 26

We first introduce some notation and facts. Let $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$ and $\phi(z) \triangleq f(z)/q_0(z) - 1$. Defining P_X as the associated PMF of μ, we can write

$$\phi(z) = \sum_x P_X(x)\frac{e^{\frac{xz}{1+x}}}{1+x} - 1, \quad (3.218)$$

which is increasing and

$$\phi(z) \geqslant \phi(0) = \mathbb{E}_\mu\left(\frac{1}{1+X}\right) - 1 \geqslant -\mathbb{E}_\mu(X) \geqslant -2\sqrt{\nu} - \nu \geqslant -0.5. \quad (3.219)$$

Furthermore, there exists a unique M_0 such that $\phi(z) \leqslant 0$ if and only if $z \leqslant M_0$.

Since we have $\log(1+x) \geqslant x - x^2/2 + \mathbb{1}\{x \leqslant 0\}\, 2x^3/3$ for $x \geqslant -0.5$, we have for all $z \in \mathcal{Z}$,

$$f(z)\log(\phi(z)+1) \geqslant f(z)\left(\phi(z) - \phi^2(z)/2\mathbb{1}\{\phi(z) \leqslant 0\}\, 2\phi(z)^3/3\right). \tag{3.220}$$

We therefore obtain

$$\begin{aligned}
\mathbb{D}(f\|q_0) &= \int_0^\infty f\log(\phi+1) && (3.221)\\
&\geqslant \int_0^M f\left(\phi - \phi^2/2\right) + \int_0^{M_0} 2f\phi^3/3 + \int_M^\infty f\log\left(f/q_0\right). && (3.222)
\end{aligned}$$

We consider each term separately in the following.

1. We have

$$\begin{aligned}
\int_0^M f\left(\phi - \phi^2/2\right) &= \int_0^M q_0(\phi+1)\left(\phi - \phi^2/2\right) && (3.223)\\
&= \int_0^M q_0\phi^2/2 + \int_0^M q_0\phi - \int_0^M q_0\phi^3/2 && (3.224)\\
&= \int_0^\infty q_0\phi^2/2 + \int_0^M q_0\phi - \int_0^M q_0\phi^3/2 - \int_M^\infty q_0\phi^2/2. && (3.225)
\end{aligned}$$

We again separately lower-bound each term in the above expression.

(a) We have by definition,

$$\int_0^\infty q_0\phi^2/2 = \frac{1}{2}\chi_2(f\|q_0). \tag{3.226}$$

(b) We have

$$\int_0^M q_0\phi = \int_0^M (f - q_0) \tag{3.227}$$

$$\geqslant -\int_M^\infty f. \tag{3.228}$$

(c) To lower-bound $-\int_0^M q_0\phi^3/2$, we first upper-bound ϕ as follows.

$$\phi(z) = \sum_x P_X(x)\frac{e^{\frac{xz}{1+x}}}{1+x} - 1 \tag{3.229}$$

$$\leqslant \sum_x P_X(x)e^{\frac{xz}{1+a}} - 1 \tag{3.230}$$

$$\overset{(a)}{\leqslant} \sum_x P_X(x)\left(1 + \left(e^{\frac{az}{1+a}} - 1\right)x\right) - 1 \tag{3.231}$$

$$= \left(e^{\frac{az}{1+a}} - 1\right)\mathbb{E}_\mu(X) \tag{3.232}$$

$$= e^{\frac{az}{1+a}}\mathbb{E}_\mu(X), \tag{3.233}$$

where (a) follows since $e^{\frac{xz}{1+a}} \leqslant 1 + \left(e^{\frac{az}{1+a}} - 1\right)x$ for $x \in [0, a]$. Since $q_0 > 0$ and $x \mapsto x^3$ is increasing, we have

$$\int_0^M q_0\phi^3 \leqslant \int_0^M e^{-z}\left(e^{\frac{az}{1+a}}\mathbb{E}_\mu(X)\right)^3 dz \tag{3.234}$$

$$= (\mathbb{E}_\mu(X))^3 \int_0^M e^{z\left(-1+\frac{3a}{1+a}\right)}dz. \tag{3.235}$$

(d) Since $\phi(z) \geqslant 0$ for $z \geqslant M \geqslant M_0$ and $x \mapsto x^2$ is increasing for $x \geqslant 0$, we have

$$\int_M^\infty q_0\phi^2 \leqslant \int_M^\infty e^{-z}\left(e^{\frac{az}{1+a}}\mathbb{E}_\mu(X)\right)^2 dz \tag{3.236}$$

$$= (\mathbb{E}_\mu(X))^2 \int_0^M e^{z\left(-1+\frac{2a}{1+a}\right)}dz. \tag{3.237}$$

As a conclusion, we obtain that

$$\int_0^M f\left(\phi - \phi^2/2\right) \geqslant \frac{1}{2}\chi_2(f\|q_0) - \frac{1}{2}(\mathbb{E}_\mu(X))^2 \int_0^M e^{z\left(-1+\frac{2a}{1+a}\right)}dz$$
$$- \frac{1}{2}(\mathbb{E}_\mu(X))^3 \int_0^M e^{z\left(-1+\frac{3a}{1+a}\right)}dz - \int_M^\infty f. \tag{3.238}$$

2. Using $|f(z)| \leqslant 1$ for all $z \in \mathcal{Z}$ and $0 \geqslant \phi(z) \geqslant -\mathbb{E}_\mu(X)$ for all $0 \leqslant z \leqslant M_0$, we have

$$\int_0^{M_0} f\phi^3 \geqslant -M_0 \left(\mathbb{E}_\mu(X)\right)^3. \tag{3.239}$$

We now show that $M_0 \leqslant 2$, for which it is enough to show that $f(2) \geqslant q_0(2)$. Note that

$$\begin{aligned}
\log f(2) &\geqslant \sum_x P_X(x) \log q_x(2) && (3.240)\\
&= -\mathbb{E}_\mu(\log(1+X)) - 2\mathbb{E}_\mu\left(\frac{1}{1+X}\right) && (3.241)\\
&\geqslant \mathbb{E}_\mu(X) - 2 + 2\mathbb{E}_\mu\left(\frac{X}{1+X}\right) && (3.242)\\
&\geqslant \mathbb{E}_\mu(X) - 2 + 2\mathbb{E}_\mu\left(\frac{X}{1+a}\right) && (3.243)\\
&\geqslant -2 = \log q_0(2). && (3.244)
\end{aligned}$$

3. By our assumption that $f(M)/q_0(M) \geqslant e$, we have

$$\int_M^\infty f \log f/q_0 \geqslant \int_M^\infty f. \tag{3.245}$$

Combining the bounds in the above three parts, we obtain the desired result.

3.E Optimization Problem in (3.18)

3.E.1 Prokhorov's Theorem

Theorem 18. *Let $\{\mu_n\}$ be a sequence of tight probability measures on $\mathbb{R}$, i.e., for all $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}$ such that for all $n \geqslant 1$, $\mu_n(\mathbb{R} \setminus K) \leqslant \epsilon$. Then, there*

exists a sub-sequence $\{\mu_{n_k}\}_{k\geqslant 1}$ *and another probability measure* μ *on* $\mathbb{R}$ *such that* $\{\mu_{n_k}\}_{k\geqslant 1}$ *converges weakly to* μ*.*

3.E.2 Convex Optimization for General Vector Spaces

Theorem 19. *([62, Theorem 1, Page 217]). Let* $\mathcal{V}$ *be a vector space,* $\Omega \subset \mathcal{V}$ *a convex set,* $\mathcal{U}$ *be a normed vector space, and* $\mathcal{P} \subset \mathcal{U}$ *be a positive cone, i.e., for all* $u_1, u_2 \in \mathcal{U}$ *and all* $\alpha, \beta \geqslant 0$*, we have* $\alpha u_1 + \beta u_2 \in \mathcal{P}$*. Suppose the interior of* $\mathcal{P}$ *is non-empty, and* $\phi : \Omega \to \mathbb{R}$ *and* $G : \Omega \to \mathcal{U}$ *are convex functions such that there exists* $\omega_1 \in \Omega$ *for which* $G(\omega_1) \prec_{\mathcal{P}} 0$ *and* $A \triangleq \inf_{\omega\in\Omega: G(\omega)\preceq_{\mathcal{P}} 0} \phi(\omega) > -\infty$*. Then, there exists* $u_0^* \succeq_{\mathcal{P}^*} 0$ *in* $\mathcal{U}^*$ *such that* $A = \inf_{\omega\in\Omega} \phi(\omega) + \langle G(\omega), u_0^*\rangle$*. Moreover, if* ω_0 *is a solution to the first optimization problem, the infimum of the second optimization problem is also achieved by* ω_0 *and* $\langle G(\omega_0), u_0^*\rangle = 0$*.*

We next recall a result from [54] to find an expression for the KKT conditions of an abstract convex optimization. To this end, we introduce the notation of weak differentiablity for a function $f : \Omega \to \mathbb{R}$ where Ω is convex. We say that $f'_{\omega_0} : \Omega \to \mathbb{R}$ is the weak derivative of f at ω_0, if

$$f'_{\omega_0}(\omega) = \lim_{\theta\to 0^+} \frac{f(\theta\omega + (1-\theta)\omega_0)}{\theta}. \tag{3.246}$$

Theorem 20 ([54])**.** *Let* $\mathcal{V}$ *be a linear space,* $\Omega \subset \mathcal{V}$ *be convex, and* $f : \Omega \to \mathbb{R}$ *be convex and have weak derivative for all* $\omega \in \Omega$*.* $f(\omega^*) = \inf_{\omega\in\Omega} f(\omega)$ *if and only if for all* $\omega \in \Omega$*, we have* $f'_{\omega^*}(\omega) \geqslant 0$*.*

3.E.3 Technical Results

Lemma 33. $A(\nu)$ *defined in* (3.18) *satisfies the following properties.*

1. *It is concave and non-decreasing on* $[0, \infty)$*.*
2. *It is continuous on* $[0, \infty)$*.*

3. *The one-sided derivatives,*

$$A'(\nu^+) \triangleq \lim_{h\to 0^+} \frac{A(\nu+h)-A(\nu)}{h} \text{ and } A'(\nu^-) \triangleq \lim_{h\to 0^+} \frac{A(\nu)-A(\nu-h)}{h}, \quad (3.247)$$

exist for all $\nu > 0$*, and for all* $0 < \nu_1 < \nu_2$*, we have* $A'(\nu_1^-) \geqslant A'(\nu_1^+) \geqslant A'(\nu_2^-) \geqslant A'(\nu_2^+)$.

4. *There exist constants* $\nu_0 > 0$ *and* $C > 0$ *such that for all* $0 < \nu \leqslant \nu_0$*, we have* $A(\nu) \geqslant C\sqrt{\nu}$.

5. *We have* $\lim_{\nu\to 0^+} A'(\nu^+) = \lim_{\nu\to 0^+} A'(\nu^-) = \infty$.

Proof. 1. By definition of $A(\nu)$, it follows that $A(\nu)$ is non-decreasing. To check concavity, we take any $\nu_1, \nu_2 > 0$, $\mu_1, \mu_2 \in \Omega$ with $\mathbb{D}\big(w_{Z|X} \circ \mu_1 \| q_0\big) \leqslant \nu_1$ and $\mathbb{D}\big(w_{Z|X} \circ \mu_2 \| q_0\big) \leqslant \nu_2$, and $\lambda \in [0,1]$. By convexity of the relative entropy, we have

$$\mathbb{D}\big(w_{Z|X} \circ (\lambda\mu_1 + (1-\lambda)\mu_2) \,\|q_0\big) \leqslant \lambda\nu_1 + (1-\lambda)\nu_2. \quad (3.248)$$

Therefore, by concavity of the mutual information,

$$A(\lambda\nu_1 + (1-\lambda)\nu_2) \geqslant I(\lambda\mu_{\nu_1}^* + (1-\lambda)\mu_{\nu_2}^*, w_{Y|X}) \quad (3.249)$$

$$\geqslant \lambda I(\mu_1, w_{Y|X}) + (1-\lambda) I(\mu_2, w_{Y|X}). \quad (3.250)$$

Hence, by definition of supremum, we have

$$A(\lambda\nu_1 + (1-\lambda)\nu_2) \quad (3.251)$$

$$\geqslant \sup_{\mu_1,\mu_2\in\Omega: \mathbb{D}\left(w_{Z|X}\circ\mu_1\|q_0\right)\leqslant\nu_1, \mathbb{D}\left(w_{Z|X}\circ\mu_2\|q_0\right)\leqslant\nu_2} \lambda I(\mu_1, w_{Y|X}) + (1-\lambda) I(\mu_2, w_{Y|X}) \quad (3.252)$$

$$= \lambda \sup_{\mu_1 \in \Omega : \mathbb{D}\left(w_{Z|X} \circ \mu_1 \| q_0\right) \leqslant \nu_1} I(\mu_1, w_{Y|X}) + (1-\lambda) \sup_{\mu_2 \in \Omega : \mathbb{D}\left(w_{Z|X} \circ \mu_2 \| q_0\right) \leqslant \nu_2} I(\mu_2, w_{Y|X}) \tag{3.253}$$

$$= \lambda A(\nu_1) + (1-\lambda) A(\nu_2). \tag{3.254}$$

2. Since $A(\nu)$ is concave on $[0, \infty)$, it is continuous on $(0, \infty)$ [63, Page 153, Problem 4]. To check the continuity at 0, we consider $\nu > 0$ and $\mu \in \Omega$ with $\mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right) \leqslant \nu$. Using (3.173), we have

$$I(\mu, w_{Y|X}) = -\int_0^\infty (w_{Y|X} \circ \mu)(y) \log (w_{Y|X} \circ \mu)(y) dy - \mathbb{E}_\mu\left(\log(1 + \theta_m^2 X)\right) - 1. \tag{3.255}$$

Furthermore, since $\mathbb{E}_{w_{Y|X} \circ \mu}(Y) = 1 + \theta_m^2 \mathbb{E}_\mu(X)$ by (3.141) and the support of $w_{Y|X} \circ \mu$ is included in $[0, \infty)$, the differential entropy of $w_{Y|X} \circ \mu$ is upper-bounded by the differential entropy of an exponential distribution with the same mean [64]. Therefore, we have

$$I(\mu, w_{Y|X}) \leqslant 1 + \log(1 + \theta_m^2 \mathbb{E}_\mu(X)) - \mathbb{E}_\mu\left(\log(1 + \theta_m^2 X)\right) - 1 \tag{3.256}$$

$$\leqslant \theta_m^2 \mathbb{E}_\mu(X). \tag{3.257}$$

Furthermore, we have

$$\nu \geqslant \mathbb{D}\left(w_{Z|X} \circ \mu \| q_0\right) \tag{3.258}$$

$$= \int_0^\infty (w_{Z|X} \circ \mu)(z) \log \frac{(w_{Z|X} \circ \mu)(z)}{q_0(z)} dz \tag{3.259}$$

$$\overset{(a)}{=} \int_0^\infty (w_{Z|X} \circ \mu)(z) \log ((w_{Z|X} \circ \mu)(z)) dz + \mathbb{E}_{w_{Z|X} \circ \mu}(Z) \tag{3.260}$$

$$\geqslant -1 - \log(\mathbb{E}_{w_{Z|X} \circ \mu}(Z)) + \mathbb{E}_{w_{Z|X} \circ \mu}(Z) \tag{3.261}$$

$$\stackrel{(b)}{=} -\log(1+\mathbb{E}_\mu(X)) + \mathbb{E}_\mu(X) \tag{3.262}$$

$$\stackrel{(c)}{\geqslant} \frac{1}{2}\mathbb{E}_\mu(X)^2 - \frac{1}{3}\mathbb{E}_\mu(X)^3 \tag{3.263}$$

$$\stackrel{(d)}{\geqslant} \frac{1}{2}\mathbb{E}_\mu(X)^2\left(1-\frac{2}{3}\left(2\sqrt{\nu}+\nu\right)\right) \tag{3.264}$$

where (a) follows since $\log q_0(z) = -z$ and $\mathbb{E}_{w_{Z|X}\circ\mu}(Z) < \infty$ by Lemma 27 and (3.142), (b) follows from (3.142), (c) follows from $\log(1+x) \leqslant x - x^2/2 + x^3/3$ for $x > -1$, and (d) follows from Lemma 28. We obtain for $\nu < 1/4$ that $\mathbb{E}_\nu(X) \leqslant \frac{\sqrt{2\nu}}{1-(2/3)(2\sqrt{\nu}+\sqrt{\nu})} \leqslant \frac{\sqrt{2\nu}}{1-2\sqrt{\nu}}$, and hence,

$$I(\mu, w_{Y|X}) \leqslant \frac{\theta_m^2\sqrt{2\nu}}{1-2\sqrt{\nu}}. \tag{3.265}$$

Additionally, since $A(\nu)$ is non-decreasing and non-negative, we have

$$|A(\nu) - A(0)| = A(\nu) - A(0) \leqslant A(\nu) \leqslant \frac{\theta_m^2\sqrt{2\nu}}{1-2\nu}, \tag{3.266}$$

which implies that $A(\nu)$ is continuous at zero.

3. Follows from [63, Page 153, Problem 4] and concavity of $A(\nu)$.

4. For $\nu > 0$ small enough, it is enough to find a probability measure μ satisfying $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) \leqslant \nu$ and $I(\mu, w_{Y|X}) \geqslant C\sqrt{\nu}$. Let μ be a discrete probability measure on $\mathcal{X}$ with two mass points at 0 and $\widetilde{x}$ with probabilities $1-\alpha$ and α, respectively, such that $\widetilde{x} < \min(1, 1/\theta_m^2)$. Then, by Lemma 32,

$$\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) = \frac{\alpha^2\widetilde{x}^2}{2(1-\widetilde{x}^2)} + o(\alpha^2). \tag{3.267}$$

Similarly, we can obtain $\mathbb{D}\big(w_{Y|X}\circ\mu\|p_0\big) \leqslant \alpha^2\theta_m^2\widetilde{x}^2/(2(1-\theta_m^2\widetilde{x}^2)) + o(\alpha^2)$. There-

fore, we can lower-bound the mutual information by

$$I(\mu, w_{Y|X}) = \alpha \mathbb{D}(p_{\widetilde{x}} \| p_0) - \mathbb{D}\big(w_{Y|X} \circ \mu \| p_0\big) \tag{3.268}$$

$$\geqslant \alpha \mathbb{D}(p_{\widetilde{x}} \| p_0) - \frac{\alpha^2 \theta_m^2 \widetilde{x}^2}{2(1 - \theta_m^2 \widetilde{x}^2)} - o(\alpha^2) \tag{3.269}$$

$$= \alpha \left(\theta_m^2 \widetilde{x} - \log(1 + \theta_m^2 \widetilde{x})\right) - \frac{\alpha^2 \theta_m^2 \widetilde{x}^2}{2(1 - \theta_m^2 \widetilde{x}^2)} - o(\alpha^2). \tag{3.270}$$

Hence, by choosing

$$\alpha = \widetilde{x}^{-1} \sqrt{2(1 - \widetilde{x}^2) \mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big)} = \widetilde{x}^{-1} \sqrt{2(1 - \widetilde{x}^2) \nu (1 - o(1))}, \tag{3.271}$$

we have $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) \leqslant \nu$ and

$$I(\mu, w_{Y|X}) \geqslant \sqrt{\nu}(1 - o(1)) \times \left(\widetilde{x}^{-1} \sqrt{2(1 - \widetilde{x}^2)} \left(\theta_m^2 \widetilde{x} - \log(1 + \theta_m^2 \widetilde{x})\right) - \sqrt{\nu} \frac{1 - \widetilde{x}^2}{1 - \theta_m^2 \widetilde{x}^2} \right). \tag{3.272}$$

Choosing $\nu_0 > 0$ such that

$$\widetilde{x}^{-1} \sqrt{2(1 - \widetilde{x}^2)} \left(\theta_m^2 \widetilde{x} - \log(1 + \theta_m^2 \widetilde{x})\right) \geqslant 2\sqrt{\nu_0} \frac{1 - \widetilde{x}^2}{1 - \theta_m^2 \widetilde{x}^2}, \tag{3.273}$$

the claim of the lemma holds for

$$C = \frac{1}{2} \widetilde{x}^{-1} \sqrt{2(1 - \widetilde{x}^2)} \left(\theta_m^2 \widetilde{x} - \log(1 + \theta_m^2 \widetilde{x})\right). \tag{3.274}$$

5. Since $A'(\nu^+) \leqslant A'(\nu^-)$, we only need to compute $\lim_{\nu \to 0^+} A'(\nu^+)$. Since $A(\nu)$ is concave $A'(\nu^+)$ is decreasing, and therefore, it is enough to show that for any $L > 0$ there exists some $\nu > 0$ with $A'(\nu^+) \geqslant L$. To this end, we fix some $\widetilde{\nu} > 0$ and define $B(\nu) \triangleq A(\nu) - \frac{A(\widetilde{\nu})}{\widetilde{\nu}} \nu$. $B(\nu)$ is continuous on $[0, \widetilde{\nu}]$ and therefore it achieves its maximum and minimum on $[0, \widetilde{\nu}]$. Hence, either we have $B(\nu) = 0$

for all $\nu \in [0, \widetilde{\nu}]$ or there exists a $\nu \in (0, \widetilde{\nu})$ such that $B(\nu)$ achieves its maximum or minimum at ν. Then, we should have $B'(\nu-) = A'(\nu^-) - A(\widetilde{\nu})/\widetilde{\nu} \geqslant 0$ or $B'(\nu^+) = A'(\nu^+) - A(\widetilde{\nu})/\widetilde{\nu} \geqslant 0$. In both cases, we have $A'\left(\frac{\nu}{2}^+\right) \geqslant \frac{A(\widetilde{\nu})}{\widetilde{\nu}}$. However, by Lemma 33, $A'\left(\frac{\nu}{2}^+\right) \geqslant C/\sqrt{\widetilde{\nu}}$, if $\widetilde{\nu} \leqslant \nu_0$. Since $\widetilde{\nu}$ is arbitrary, we can choose it such that $C/\sqrt{\widetilde{\nu}} > L$.

□

Proof of Lemma 18. We only prove the existence of a solution and the uniqueness follows from strict concavity of the mutual information [56]. Consider a sequence $\{\mu_n\}_{n\geqslant 1}$ in Ω such that $\mathbb{D}\big(w_{Z|X} \circ \mu_n \| q_0\big) \leqslant \nu$ and $\lim_{n\to\infty} I(\mu_n, w_{Y|X}) = A(\nu)$. To use 18, we first check that this sequence is tight. For any $\epsilon > 0$, we have

$$\mathbb{P}_{\mu_n}\big(X \notin [0, (2\sqrt{\nu}+\nu)/\epsilon]\big) \overset{(a)}{\leqslant} \frac{\mathbb{E}_{\mu_n}(X)\epsilon}{2\sqrt{\nu}+\nu} \tag{3.275}$$

$$\overset{(b)}{\leqslant} \epsilon, \tag{3.276}$$

where (a) follows from applying Markov's inequality to the almost surely non-negative random variable X, and (b) follows from Lemma 28. Since $[0, (2\sqrt{\nu}+\nu)/\epsilon]$ is compact, the sequence $\{\mu_n\}_{n\geqslant 1}$ is tight. Therefore, we are permitted to use Theorem 18 that shows the existence of a subsequence $\{\mu_{n_k}\}_{k\geqslant 1}$ and probability measure μ on $\mathbb{R}$ such that $\{\mu_{n_k}\}_{k\geqslant 1}$ converges weakly to μ. We claim that μ_ν^* is indeed μ and prove it in three steps.

Step 1 Theorem 18 only guarantees the existence of a probability measure on $\mathbb{R}$ which can possibly have positive measure on negative numbers. In this step, we show that this is not the case. By the Portmanteau theorem, the weak convergence of $\{\mu_{n_k}\}_{k\geqslant 1}$ to μ implies that $\liminf_{k\to\infty} \mu_{n_k}(U) \geqslant \mu(U)$ for any open set $U \subset \mathbb{R}$. Taking $U =]-\infty, 0[$, we obtain that

$$0 = \liminf_{k\to\infty} \mu_{n_k}(]-\infty, 0[) \geqslant \mu(]-\infty, 0[) \geqslant 0, \tag{3.277}$$

which means that $\mu(]-\infty, 0[) = 0$.

Step 2 In this step we prove that μ satisfies the optimization constraint, i.e., $\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) \leqslant \nu$. Let us define $f_k(z) \triangleq (w_{Z|X} \circ \mu_{n_k})(z)$ and $f(z) \triangleq (w_{Z|X} \circ \mu)(z)$. Since for any $z \in \mathcal{Z}$, $q_x(z) = e^{-z/(1+x)}/(1+x)$ is a continuous and bounded function in x, by weak convergence definition, we have

$$f_k(z) = \mathbb{E}_{\mu_{n_k}}(q_X(z)) \to \mathbb{E}_{\mu}(q_X(z)) = f(z). \tag{3.278}$$

In the next lemma, we show that $|f_k(z) \log f_k(z)|$ is uniformly upper-bounded by an integrable function.

Lemma 34. *There exists some $\widetilde{z}$ such that for all k,*

$$|f_k(z) \log f_k(z)| \leqslant g(z) \triangleq \begin{cases} e^{-1} & z \in [0, \widetilde{z}], \\ \frac{2\sqrt{\nu}+\nu}{e(z^{\frac{3}{2}} - z^{\frac{1}{2}})} + z^{-\frac{1}{2}} e^{-\sqrt{z}} & z \in [\widetilde{z}, \infty[, \end{cases} \tag{3.279}$$

and $\int_0^\infty |g(z)| dz < \infty$.

Proof. Note first that for all $x \in [0,1]$, we have $|x \log x| \leqslant e^{-1}$, and for all $x \in [0, e^{-1}]$, we have $|x \log x| \leqslant |x|$. Thus, it is enough to show that there exist $\widetilde{z}$ such that for all $k \geqslant 1$ and $z \geqslant \widetilde{z}$,

$$f_k(z) \leqslant \frac{2\sqrt{\nu}+\nu}{e(z^{\frac{3}{2}} - z^{\frac{1}{2}})} + z^{-\frac{1}{2}} e^{-\sqrt{z}}. \tag{3.280}$$

By law of total probability, for all $\lambda > 0$, we have

$$\begin{aligned} f_k(z) &\triangleq \mathbb{E}_{\mu_{n_k}}(q_X(z)) && (3.281)\\ &= \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geqslant \lambda)\mathbb{P}_{\mu_{n_k}}(X \geqslant \lambda) + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda)\mathbb{P}_{\mu_{n_k}}(X < \lambda) && (3.282)\\ &\overset{(a)}{\leqslant} \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geqslant \lambda)\frac{\mathbb{E}_{\mu_{n_k}}(X)}{\lambda} + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda) && (3.283) \end{aligned}$$

$$\overset{(b)}{\leqslant} \mathbb{E}_{\mu_{n_k}}(q_X(z)|X \geqslant \lambda)\frac{2\sqrt{\nu}+\nu}{\lambda} + \mathbb{E}_{\mu_{n_k}}(q_X(z)|X < \lambda), \tag{3.284}$$

where (a) follows from Markov's inequality, and (b) follows from Lemma 28. We also have for all $z \geqslant 1$, $q_x(z) \leqslant (ze)^{-1}$, and for all $0 \leqslant x \leqslant \lambda \leqslant z-1$, $q_x(z) \leqslant e^{-\frac{z}{1+\lambda}}/(1+\lambda)$. Substituting these upper-bounds in (3.284) for $\lambda = z^{\frac{1}{2}} - 1$, which is less than $z-1$ for $z \geqslant 1$, we obtain

$$f_k(z) \leqslant \frac{1}{ze}\frac{2\sqrt{\nu}+\nu}{z^{\frac{1}{2}}-1} + \frac{1}{z^{\frac{1}{2}}}e^{-z^{\frac{1}{2}}} \tag{3.285}$$

$$= \frac{2\sqrt{\nu}+\nu}{e(z^{\frac{3}{2}}-z^{\frac{1}{2}})} + z^{-\frac{1}{2}}e^{-\sqrt{z}}. \tag{3.286}$$

□

We are now eligible to use dominated convergence theorem and exchange limit and integral to obtain

$$\lim_{k\to\infty}\int_0^\infty f_k(z)\log f_k(z)dz = \int_0^\infty \lim_{k\to\infty} f_k(z)\log f_k(z)dz \tag{3.287}$$

$$= \int_0^\infty f(z)\log f(z)dz. \tag{3.288}$$

Since $f_k(z)z \geqslant 0$ for all $z \in \mathcal{Z}$ and $k \geqslant 1$, Fatou's lemma yields that

$$\int_0^\infty f(z)zdz = \int_0^\infty \liminf_{k\to\infty} f_k(z)zdz \tag{3.289}$$

$$\leqslant \liminf_{k\to\infty}\int_0^\infty f_k(z)zdz. \tag{3.290}$$

Combing (3.288) and (3.290), we have

$$\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) = \int_0^\infty f(z)\left(z+\log f(z)\right)dz \tag{3.291}$$

$$\leqslant \liminf_{k\to\infty} \int_0^\infty f_k(z)\,(z+\log f_k(z))\,dz = \liminf_{k\to\infty} \mathbb{D}\big(w_{Z|X}\circ\mu_{n_k}\|q_0\big) \leqslant \nu. \tag{3.292}$$

Step 3 It remains to show that $I(\mu, w_{Y|X}) \geqslant A(\nu)$. We again define $h_k(z) \triangleq (w_{Y|X}\circ\mu_{n_k})(z)$ and $h(z) \triangleq (w_{Y|X}\circ\mu)(z)$. With the same argument of the previous step, we can prove that

$$\lim_{k\to\infty} \int_0^\infty h_k(z)\log h_k(z)dz = \int_0^\infty h(z)\log h(z)dz. \tag{3.293}$$

Furthermore, by [65, Page 86], we have

$$\liminf_{k\to\infty} \mathbb{E}_{\mu_{n_k}}\big(1+\log(1+\theta_m^2 X)\big) \geqslant \mathbb{E}_{\mu}\big(1+\log(1+\theta_m^2 X)\big). \tag{3.294}$$

Hence, (3.173) implies that $I(\mu, w_{Y|X}) \geqslant A(\nu)$. □

Proof of Theorem 14. We prove all four statements in order. The proof heavily relies on results from convex optimization for general vector spaces and properties of the optimization problem in (3.19), which we have gathered in Appendix 3.E for the reader's convenience.

1. In Theorem 19, taking Ω as the set of all probability measures μ on $\mathcal{X}$ with $\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) < \infty$, $\mathcal{U} = \mathbb{R}$, $\mathcal{P} = \mathbb{R}^+$, $\phi(\mu) = -I(\mu, w_{Y|X})$, $G(\mu) = \mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu$, we note that

$$-\infty < -A(\nu) = -\sup_{\mu\in\Omega: G(\mu)\leqslant 0} -\phi(\mu) \tag{3.295}$$

$$= \inf_{\mu\in\Omega: G(\mu)\leqslant 0} \phi(\mu). \tag{3.296}$$

By convexity of the relative entropy and concavity of mutual information in the input distribution, ϕ and G are convex functions, with μ_1 the deterministic probability measure with all mass point at zero, we also have $G(\mu_1) = -\nu < 0$. Therefore, we

can apply Theorem 19 to show the existence of $\gamma(\nu) \geqslant 0$ such that

$$\inf_{\mu\in\Omega:G(\mu)\leqslant 0} \phi(\mu) = \inf_{\mu\in\Omega} \left[\phi(\mu) + \gamma(\nu)G(\mu)\right] \tag{3.297}$$

$$= -\sup_{\mu\in\Omega} \left[I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu\right)\right], \tag{3.298}$$

which results in the unconstrained reformulation of $A(\nu)$ as

$$\sup_{\mu\in\Omega} \left[I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu\right)\right]. \tag{3.299}$$

Theorem 19 also implies that μ_ν^* is a solution to this new optimization problem, and since $I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu\right)$ is strictly concave [56, Appendix I.B], the solution is unique.

2. With the help of Lemma 35 in Appendix C to show the existence of weak derivatives (defined in (3.246)), we use Theorem 20 with

$$f(\mu) = I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu\right) \tag{3.300}$$

to obtain that $\mu_1 = \mu_\nu^*$ if and only if for any $\mu \in \Omega$,

$$0 \geqslant f'_{\mu_1}(\mu) \tag{3.301}$$

$$= \mathbb{E}_{w_{Y|X}\otimes\mu}\left(\log\frac{p_X(Y)}{(w_{Y|X}\circ\mu_1)(Y)}\right) - I(\mu_1, w_{Y|X}) - \gamma(\nu)\left(\mathbb{E}_{w_{Z|X}\circ\mu}\left(\log\frac{(w_{Z|X}\circ\mu_1)(Z)}{q_0(Z)}\right) - \mathbb{D}\big(w_{Z|X}\circ\mu_1\|q_0\big)\right) \tag{3.302}$$

$$= \mathbb{E}_{w_{Y|X}\otimes\mu}\left(\log\frac{p_X(Y)}{(w_{Y|X}\circ\mu_1)(Y)}\right) - \gamma(\nu)\left(\mathbb{E}_{w_{Z|X}\circ\mu}\left(\log\frac{(w_{Z|X}\circ\mu_1)(Z)}{q_0(Z)}\right) - \nu\right) - f(\mu_1) \tag{3.303}$$

$$= \mathbb{E}_\mu(w(X, \mu_1, \nu)) - f(\mu_1). \tag{3.304}$$

This implies that $\mu_1 = \mu_\nu^*$ if and only if for all $\mu \in \Omega$, we have $f(\mu_1) \geqslant \mathbb{E}_\mu(w(X, \mu_1, \nu))$. Since $A(\nu) = \sup_{\mu\in\Omega} f(\mu) \geqslant f(\mu_1)$, if $\mu_1 = \mu_\nu^*$, then for all $\mu \in \Omega$, we have $A(\nu) \geqslant f(\mu_1) \geqslant \mathbb{E}_\mu(w(X, \mu_1, \nu))$.

3. Assume (3.23) is true, we take the expectation and obtain (3.21). We now show the opposite direction and prove that if (3.21) holds, we have (3.23) and (3.24). Applying (3.21) with μ a deterministic probability measure with all mass point at x, we obtain

$$A(\nu) \geqslant \mathbb{E}_\mu(w(X, \mu_1, \nu)) = w(x, \mu_1, \nu). \tag{3.305}$$

Furthermore, for any $x \in \text{support}\,(\mu_1)$, we prove that $w(x, \mu_1, \nu) = A(\nu)$ by contradiction. If $A(\nu) - w(x, \mu_1, \nu) \triangleq \delta > 0$, by continuity of $A(\nu) - w(x, \mu_1, \nu)$ in x, there exists a neighborhood $\mathcal{N}$ of x such that for all $x' \in \mathcal{N}$, we have $A(\nu) - w(x', \mu_1, \nu) \geqslant \delta/2$. Also, since $x \in \text{support}\,(\mu_1)$, we know that $\mathbb{P}_{\mu_1}(X \in \mathcal{N}) = \epsilon > 0$. Therefore, we obtain

$$\begin{aligned}
A(\nu) &= \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)) && (3.306)\\
&= \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)\mathbb{1}\{X \in \mathcal{N}\}) + \mathbb{E}_{\mu_1}(w(X, \mu_1, \nu)\mathbb{1}\{X \notin \mathcal{N}\}) && (3.307)\\
&\leqslant (1 - \epsilon)A(\nu) + \epsilon\left(A(\nu) - \frac{\delta}{2}\right) && (3.308)\\
&= A(\nu) - \frac{\delta\epsilon}{2} < A(\nu), && (3.309)
\end{aligned}$$

which is a contradiction.

4. To prove that $\lim_{\nu\to 0^+} \gamma(\nu) = \infty$, we prove that $\gamma(\nu) \geqslant A'(\nu^+)$, and the result will follow from $\lim_{\nu\to 0^+} A'(\nu^+) = \infty$ as shown in Lemma 33. Consider any $\nu_1, \nu_2 > 0$,

and similar to the sensitivity analysis in [57, Section 5.6], note that

$$A(\nu_1) = I(\mu^*_{\nu_1}, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}\big(w_{Z|X} \circ \mu^*_{\nu_1} \| q_0\big) - \nu_1) \tag{3.310}$$

$$\overset{(a)}{\geqslant} I(\mu^*_{\nu_2}, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}\big(w_{Z|X} \circ \mu^*_{\nu_2} \| q_0\big) - \nu_1) \tag{3.311}$$

$$= I(\mu^*_{\nu_2}, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}\big(w_{Z|X} \circ \mu^*_{\nu_2} \| q_0\big) - \nu_2) + \gamma(\nu_1)(\nu_1 - \nu_2) \tag{3.312}$$

$$\overset{(b)}{\geqslant} I(\mu^*_{\nu_2}, w_{Y|X}) + \gamma(\nu_1)(\nu_1 - \nu_2), \tag{3.313}$$

where (a) follows since $\mu^*_{\nu_1}$ is the maximizer of

$$\sup_{\mu} I(\mu, w_{Y|X}) - \gamma(\nu_1)(\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) - \nu_1), \tag{3.314}$$

and (b) follows since $\gamma(\nu_1) \geqslant 0$. Thus, for any $\nu > 0$ and $\nu > h > 0$, we have

$$\frac{A(\nu) - A(\nu - h)}{h} \geqslant \gamma(\nu) \text{ and } \frac{A(\nu + h) - A(\nu)}{h} \leqslant \gamma(\nu). \tag{3.315}$$

Taking the limit $h \to 0^+$, we obtain $A'(\nu^+) \leqslant \gamma(\nu) \leqslant A'(\nu^-)$.

To prove that $\lim_{\nu \to 0^+} \gamma(\nu)\nu = 0$, note that for all $\nu > 0$,

$$\gamma(\nu)\nu \leqslant A'(\nu^-)\nu \tag{3.316}$$

$$\overset{(a)}{\leqslant} \frac{A(\nu)}{\nu}\nu = A(\nu), \tag{3.317}$$

where (a) follows from concavity of A. In the proof of Lemma 33, we show that $\lim_{\nu \to 0^+} A(\nu) = 0$, which yields the result.

□

Lemma 35. *$f(\mu) \triangleq I(\mu, w_{Y|X}) - \gamma(\nu)\left(\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) - \nu\right)$ is weakly differentiable,*

and

$$f'_{\mu_1}(\mu) = \mathbb{E}_{w_{Y|X}\otimes\mu}\left(\log\frac{p_X(Y)}{(w_{Y|X}\circ\mu_1)(Y)}\right) - I(\mu_1, w_{Y|X}) - \gamma(\nu)\left(\mathbb{E}_{w_{Z|X}\circ\mu}\left(\log\frac{(w_{Z|X}\circ\mu_1)(Z)}{q_0(Z)}\right) - \mathbb{D}\big(w_{Z|X}\circ\mu_1\|q_0\big)\right). \quad (3.318)$$

Proof. In [56, Equation (63)], the weak derivative of $I(\mu, w_{Y|X})$ at μ_1 is proved to be

$$\mathbb{E}_{w_{Y|X}\otimes\mu}\left(\log\frac{p_X(Y)}{(w_{Y|X}\circ\mu_1)(Y)}\right) - I(\mu_1, w_{Y|X}). \quad (3.319)$$

Thus, we only check the weak differentiability of $G(\mu) \triangleq \mathbb{D}\big(w_{Z|X}\circ\mu\|q_0\big) - \nu$. Let $\mu_1, \mu \in \Omega$, and define

$$\mu_\theta \triangleq (1-\theta)\mu_1 + \theta\mu \quad (3.320)$$

$$f_1(z) \triangleq (w_{Z|X}\circ\mu_1)(z) \quad (3.321)$$

$$f(z) \triangleq (w_{Z|X}\circ\mu)(z) \quad (3.322)$$

$$f_\theta(z) \triangleq (w_{Z|X}\circ\mu_\theta)(z). \quad (3.323)$$

Then, we have

$$G(\mu_\theta) - G(\mu_1) \quad (3.324)$$

$$= \mathbb{D}(f_\theta\|q_0) - \mathbb{D}(f_1\|q_0) \quad (3.325)$$

$$= \int_0^\infty f_\theta(z)\log\frac{f_\theta(z)}{q_0(z)}dz - \int_0^\infty f_1(z)\log\frac{f_1(z)}{q_0(z)}dz \quad (3.326)$$

$$\overset{(a)}{=} \int_0^\infty f_\theta(z)\log\frac{f_\theta(z)}{q_0(z)}dz - \int_0^\infty f_\theta(z)\log\frac{f_1(z)}{q_0(z)}dz + \int_0^\infty f_\theta(z)\log\frac{f_1(z)}{q_0(z)}dz - \int_0^\infty f_1(z)\log\frac{f_1(z)}{q_0(z)}dz \quad (3.327)$$

$$= \int_0^\infty f_\theta(z)\log\frac{f_\theta(z)}{f_1(z)}dz + \int_0^\infty f_\theta(z)\log\frac{f_1(z)}{q_0(z)}dz - \int_0^\infty f_1(z)\log\frac{f_1(z)}{q_0(z)}dz \quad (3.328)$$

$$\overset{(b)}{=} \int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} dz + \theta \left(\int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz \right), \tag{3.329}$$

where (a) holds since by Lemma 30, $\int_0^\infty f_\theta(z) \log \frac{f_1(z)}{q_0(z)} dz < \infty$, and (b) follows from $f_\theta = (1-\theta) f_1 + \theta f$. The second term in (3.329) is differentiable with respect to θ, and the derivative is

$$\int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz. \tag{3.330}$$

To take derivative from the first term in (3.329), we use Theorem 16. Note that by Lemma 30, $\int_0^\infty f_\theta(z) \left| \log \frac{f_\theta(z)}{f_1(z)} \right| dz < \infty$, and also, for all z and θ,

$$\frac{\partial}{\partial \theta} \left(f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} \right) = -(f_1(z) - f(z)) \left(1 + \log \frac{f_\theta(z)}{f_1(z)} \right). \tag{3.331}$$

Additionally, for all $\theta \in [0,1]$, if we apply (3.140), we obtain

$$\left| f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} \right| \leqslant |f_1(z) + f(z)| \left(|\log f_1(z)| + \log(1 + \mathbb{E}_{\mu_\theta}(X)) + z \right) \tag{3.332}$$

$$\leqslant |f_1(z) + f(z)| \left(|\log f_1(z)| + \log(1 + \mathbb{E}_{\mu_1}(X) + \mathbb{E}_{\mu}(X)) + z \right), \tag{3.333}$$

which is a integrable function with respect to Lebesgue measure on $\mathcal{Z}$ by Lemma 30 and does not depend on θ. Hence, all condition in Theorem 16 hold, and we have

$$\frac{\partial}{\partial \theta} \left(\int_0^\infty f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} dz \right) = \int_0^\infty \frac{\partial}{\partial \theta} \left(f_\theta(z) \log \frac{f_\theta(z)}{f_1(z)} \right) dz \tag{3.334}$$

$$= \int_0^\infty -(f_1(z) - f(z)) \left(1 + \log \frac{f_\theta(z)}{f_1(z)} \right) dz \tag{3.335}$$

$$= \int_0^\infty -(f_1(z) - f(z)) \log \frac{f_\theta(z)}{f_1(z)} dz, \tag{3.336}$$

which vanishes at $\theta = 0$. Therefore, G is weakly differentiable at μ_1 and

$$G'_{\mu_1}(\mu) = \int_0^\infty f(z) \log \frac{f_1(z)}{q_0(z)} dz - \int_0^\infty f_1(z) \log \frac{f_1(z)}{q_0(z)} dz. \tag{3.337}$$

Since the mutual information and the divergence are weakly differentiable, so is $I(\mu, w_{Y|X}) - \gamma(\nu) \big(\mathbb{D}\big(w_{Z|X} \circ \mu \| q_0\big) - \nu \big)$. □

CHAPTER 4
FUNDAMENTAL LIMITS OF COVERT COMMUNICATION: THE KNOWLEDGE OF WARDEN ABOUT THE CODE

4.1 Summary

We investigate here how the warden's knowledge of the code used by the transmitter and the receiver influences the optimal number of bits that can be covertly and reliably sent. We formulate the problem in three scenarios: the warden completely knows the code, the warden knows only the rate and an upper-bound on the probability of error at the decoder, and the warden has access to previous samples of its channel when the code was used. The first scenario corresponds to the standard model in covert communication. For the second scenario, we characterize the number of bits that can be transmitted reliably subject to a covertness constraint up to the first-order of asymptotics. Finally, for the third scenario, we provide lower- and upper-bound on the number of bits that can be transmitted reliably subject to a covertness constraint for two scalings of the number of observed samples. The content of this chapter is based on [66].

4.2 Introduction

Virtually all results on covert communications highly depend on two premises in the problem formulation, which we review in details next: how to measure covertness and which knowledge the parties have about the parameters of the underlying channel statistical model. In covert communication problems, it is commonly assumed that there exists a particular channel input symbol that can be considered as no communication (e.g., the transmission of zero over an AWGN channel); the covertness is then defined in terms of the distance between the distribution of the warden's observations when no communication happens and

when transmission happens. The frequently used distance is the relative entropy [12, 13], which can be shown to control the performance of the warden's detector through Pinsker's inequality. Another perhaps more operational distance is the total variation [39], which more directly controls the performance of the warden's detection. It is also generally assumed that all parties are aware of the channels statistics and the code structure. While this assumption seems reasonable from the perspective of the two legitimate parties who aim at communicating reliably, it corresponds to a *worst case* scenario in which the warden is granted as much prior information as the legitimate parties. The fact that the warden completely knows the code is implicitly assumed in the definition of the covertness metric, which depends on the induced distribution when the code is used.

We address here the problem of covert communication when the warden does not know the complete structure of the code. Since the induced distribution on the warden's observations by the code is only meaningful when the warden knows the code, alternative covertness metrics have to be developed depending on the assumptions. We first assume that the warden only knows the rate and an upper-bound on the probability of error, and desires to devise a test that performs well for all codes satisfying such constraints. Under this scenario, we characterize the number of bits that can be transmitted covertly and reliably. In particular, we show that the number of bits, up to first order asymptotics, is equal to the number of bits when the transmitter and the receiver have access to a shared source of randomness and the code is known to the receiver. We then consider a middle ground between the assumptions of completely known code and completely unknown code, in which the warden is allowed to query samples from its channel output when the code is used, which is then used as side information to detect a subsequent transmission. We derive asymptotically close lower- and upper-bounds for the required number of samples to detect the communication with low probability of error.

4.3 Problem Formulation

We recall the basic setup for covert communication from Section 1.1.1. Let the transmitter (Alice) be connected to the receiver (Bob) and the warden (Willie) through two binary-input DMCs $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ and $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$, respectively. We assume that $\mathcal{X} = \{0, 1\}$, in which 0 is the symbol used in the absence of communication. Let $P_x \triangleq W_{Y|X=x}$, $Q_x \triangleq W_{Z|X=x}$ for $x \in \mathcal{X}$ and $P^{\mathbf{x}} \triangleq P_{x_1} \otimes \cdots \otimes P_{x_n}$, $Q^{\mathbf{x}} \triangleq Q_{x_1} \otimes \cdots \otimes Q_{x_n}$ for $\mathbf{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$. We assume that $Q_1 \ll Q_0$ and $P_1 \ll P_0$, and $Q_1 \neq Q_0$ and define $\mu_0 \triangleq \min_{z:Q_1(z)>0} Q_0(z)$. Let $\mathbf{0}$ denote the all-zero sequence in $\mathcal{X}^n$, and in particular, let $Q^{\mathbf{0}} = Q_0^{\otimes n}$ be the distribution that Willie expects when there is no communication. A code $\mathcal{C}$ with a message set $\mathcal{M}$ and a block-length n consists of an encoder $f : \mathcal{M} \to \mathcal{X}^n$ and a decoder $\phi : \mathcal{Y}^n \to \mathcal{M}$. The code $\mathcal{C}$ is called an (M, ϵ) code if $|\mathcal{M}| = M$ and

$$P^{f(m)}(\phi^{-1}(m)) \geqslant 1 - \epsilon, \forall m \in \mathcal{M}. \tag{4.1}$$

Let $\widehat{Q} \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} Q^{f(m)}$ be the distribution induced over $\mathcal{Z}^n$ when using the code $\mathcal{C} = (f, \phi)$. Willie applies a test $t : \mathcal{Z}^n \to \{0, 1\}$ to his received sequence to detect the communication. We define the probability of false alarm and missed detection as

$$\alpha(t) \triangleq Q^{\mathbf{0}}(t^{-1}(1)) \tag{4.2}$$

$$\beta(t; f) \triangleq \widehat{Q}(t^{-1}(0)), \tag{4.3}$$

A randomized code $C = (F, \Phi)$ with a message set $\mathcal{M}$ and a block-length n is a random variable with realizations in the set of all possible codes with message set $\mathcal{M}$ and block-length n. It is called an $(M, \epsilon)_{\mathrm{R}}$ code if it is an (M, ϵ) code almost surely. We define for a randomized code $C = (F, \Phi)$, $\widehat{Q} \triangleq \sum_f \mathbb{P}(F = f) \frac{1}{M} \sum_{m \in \mathcal{M}} Q^{f(m)}$ and $\beta(t; F) \triangleq \widehat{Q}(t^{-1}(0))$, in which it is implicitly assumed that the randomness of the code is kept secret from the warden. We consider three scenarios for covert communication based on Willie's

knowledge of the code.

Scenario One We assume that Willie completely knows the code, which could be deterministic or random. We then characterize the optimal performance by

$$\delta^*(M,\epsilon) \triangleq 1 - \sup_{\mathcal{C}:(M,\epsilon)} \inf_t \left[\alpha(t) + \beta(t;f)\right] \tag{4.4}$$

$$= \inf_{\mathcal{C}:(M,\epsilon)} \frac{1}{2}\left\|\widehat{Q} - Q^{\mathbf{0}}\right\|_1, \tag{4.5}$$

$$M^*(\epsilon,\delta) \triangleq \sup\{M : \delta^*(M,\epsilon) \leqslant \delta\}, \tag{4.6}$$

where in (4.4) the optimization of the test is with respect to a specific code. We similarly define $\delta_R^*(M,\epsilon)$ and $M_R^*(\epsilon,\delta)$ by considering random codes instead of deterministic codes.

Scenario Two We now consider a situation, in which Willie is unaware of the code and should find a test with low probability of error against all possible (M,ϵ) codes. This leads to swapping the $\inf$ and $\sup$ in (4.4) and results in

$$\overline{\delta}^*(M,\epsilon) \triangleq 1 - \inf_t \sup_{\mathcal{C}:(M,\epsilon)} \left[\alpha(t) + \beta(t;f)\right], \tag{4.7}$$

$$\overline{M}^*(\epsilon,\delta) \triangleq \sup\{M : \overline{\delta}^*(M,\epsilon) \leqslant \delta\}. \tag{4.8}$$

Scenario Three We propose here an intermediate ground for Willie's knowledge about the code, in which Willie can learn the code from previous observations. Alice and Bob fix an (M,ϵ) code $\mathcal{C}$ to covertly communicate. Willie has access to ℓ independent samples $\mathbf{z}^{(1)},\cdots,\mathbf{z}^{(\ell)}$ of $\widehat{Q}$ and aims at testing whether his new observations $\mathbf{z}$ comes from $\widehat{Q}$ or $Q^{\mathbf{0}}$. Willie uses a test $T : (\mathcal{Z}^n)^\ell \times \mathcal{Z}^n \to \{0,1\}$, for which $T(\mathbf{z}^{(1)},\cdots,\mathbf{z}^{(\ell)},\mathbf{z}) = 0$ indicates

no communication.

$$\alpha_\ell(T;f) \triangleq (\widehat{Q}^{\otimes \ell} \otimes Q^{\mathbf{0}})(T^{-1}(1)) \tag{4.9}$$

$$\beta_\ell(T;f) \triangleq \widehat{Q}^{\otimes \ell+1}(T^{-1}(0)). \tag{4.10}$$

We then investigate the quantities

$$\overline{\delta}^*_\ell(M,\epsilon) \triangleq 1 - \inf_{T} \sup_{\mathcal{C}:(M,\epsilon)} [\alpha_\ell(T;f) + \beta_\ell(T;f)] \tag{4.11}$$

$$\overline{M}^*_\ell(\epsilon,\delta) \triangleq \sup\{M : \overline{\delta}^*_\ell(M,\epsilon) \leqslant \delta\}. \tag{4.12}$$

Note that our definitions imply that

$$M^*(\epsilon,\delta) \leqslant M^*_R(\epsilon,\delta) \leqslant \overline{M}^*(\epsilon,\delta). \tag{4.13}$$

and for $\ell_1 \leqslant \ell_2$,

$$M^*(\epsilon,\delta) \leqslant \overline{M}^*_{\ell_2}(\epsilon,\delta) \leqslant \overline{M}^*_{\ell_1}(\epsilon,\delta) \leqslant \overline{M}^*_0(\epsilon,\delta) = \overline{M}^*(\epsilon,\delta). \tag{4.14}$$

Remark 6. *In the most general setting, one should consider local randomness at the encoder. However, we restrict the results of the paper to deterministic encoders for simplicity.*

4.4 Main Results

4.4.1 Covert Communication without Learning

In this section, we characterize $\log M^*(\epsilon,\delta)$, $\log M^*_R(\epsilon,\delta)$, and $\log \overline{M}^*(\epsilon,\delta)$ up to the first-order asymptotics. Note that the first order of $\log M^*(\epsilon,\delta)$ when $\mathbb{D}(Q_1\|Q_0) < \mathbb{D}(P_1\|P_0)$ and the first order of $\log M^*_R(\epsilon,\delta)$ have already been established in Theorem 8 in Chapter 2.

Theorem 21. *Suppose that* $\mathbb{D}(Q_1\|Q_0) > \mathbb{D}(P_1\|P_0)$. *We have*

$$\log M^*(\epsilon,\delta) = O(1), \tag{4.15}$$

$$Cn^{\frac{1}{2}} - O\left(n^{\frac{1}{4}}\right) \leqslant \log M_R^*(\epsilon,\delta) \leqslant \log \overline{M}^*(\epsilon,\delta) \leqslant Cn^{\frac{1}{2}} - \Omega\left(n^{\frac{1}{4}}\right), \tag{4.16}$$

where $C = \frac{2\mathbb{Q}\left(\frac{1-\delta}{2}\right)\mathbb{D}(P_1\|P_0)}{\sqrt{\chi_2(Q_1\|Q_0)}}$.

Suppose that $\mathbb{D}(Q_1\|Q_0) < \mathbb{D}(P_1\|P_0)$. *We have*

$$Cn^{\frac{1}{2}} - O\left(n^{\frac{1}{4}}\right) \leqslant \log M^*(\epsilon,\delta) \leqslant \log M_R^*(\epsilon,\delta) \leqslant \log \overline{M}^*(\epsilon,\delta) \leqslant Cn^{\frac{1}{2}} - \Omega\left(n^{\frac{1}{4}}\right) \tag{4.17}$$

Remark 7. *The constant hidden in* $O\left(n^{\frac{1}{4}}\right)$ *in* (4.16) *and* (4.17) *should be larger than the constant hidden in* $\Omega\left(n^{\frac{1}{4}}\right)$. *Otherwise the inequalities are not feasible. Additionally, the dependence of* $\log M^*(\epsilon,\delta)$, $\log M_R^*(\epsilon,\delta)$, *and* $\log \overline{M}^*(\epsilon,\delta)$ *on* ϵ *is hidden in the terms* $\Omega\left(n^{\frac{1}{4}}\right)$ *and* $O\left(n^{\frac{1}{4}}\right)$.

Achievability proof of Theorem 21. We have $Cn^{\frac{1}{2}} - O\left(n^{\frac{1}{4}}\right) \leqslant \log M_R^*(\epsilon,\delta)$ and $Cn^{\frac{1}{2}} - O\left(n^{\frac{1}{4}}\right) \leqslant \log M^*(\epsilon,\delta)$ by Theorem 8 in Chapter 2. □

Remark 8. *A shared key is required in the achievability proof when* $\mathbb{D}(Q_1\|Q_0) > \mathbb{D}(P_1\|P_0)$ *because the output statistics of Willie's channel are controlled through channel resolvability techniques [13], which require randomness of size* $\approx \mathbb{D}(Q_1\|Q_0)n^{\frac{1}{2}}$. *However, the randomness in the message itself is proportional to* $\mathbb{D}(P_1\|P_0)n^{\frac{1}{2}} < \mathbb{D}(Q_1\|Q_0)n^{\frac{1}{2}}$.

Converse proof of Theorem 21. We first establish a general converse for covert communication. We recall the following definition regarding universal hypothesis testing. Consider

two families of probability distributions $\mathcal{P}$ and $\mathcal{Q}$ over the same set and define

$$\beta_\alpha(\mathcal{P}, \mathcal{Q}) \triangleq \inf_{A: P(A) \leqslant \alpha, \forall P \in \mathcal{P}} \sup_{Q \in \mathcal{Q}} (1 - Q(A)). \tag{4.18}$$

In other words, $\beta_\alpha(\mathcal{P}, \mathcal{Q})$ is the minimum number β such that there exists a test with probability of false alarm and missed detection less than α and β, respectively, regardless of the choice of $(P, Q) \in \mathcal{P} \times \mathcal{Q}$.

Lemma 36. *We define for any $\mathcal{S} \subset \mathcal{X}^n$*

$$\overline{\lambda}^*(\mathcal{S}, \delta) \triangleq 1 - \delta - \inf_{\alpha \in [0,1]} \left[\alpha + \beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x} \in \mathcal{S}})\right], \tag{4.19}$$

$$\lambda^*(\mathcal{S}, \delta, M) \triangleq 1 - \delta - \sup_{\mathcal{A} \subset \mathcal{S}: |\mathcal{A}| \leqslant M} \inf_{\alpha \in [0,1]} \left[\alpha + \beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x} \in \mathcal{A}})\right]. \tag{4.20}$$

We then have

$$\overline{M}^*(\epsilon, \delta) \leqslant \inf_{\mathcal{S} \subset \mathcal{X}^n: \overline{\lambda}^*(\mathcal{S}, \delta) > 0} \frac{1}{\overline{\lambda}^*(\mathcal{S}, \delta)} \sup_{\mathbf{x} \in \mathcal{X}^n \setminus \mathcal{S}} \frac{1}{\beta_\epsilon(P^{\mathbf{x}}, P^{\mathbf{0}})}, \tag{4.21}$$

$$M^*(\epsilon, \delta) \leqslant \inf_{\mathcal{S} \subset \mathcal{X}^n: \lambda^*(\mathcal{S}, \delta, M^*(\epsilon, \delta)) > 0} \frac{1}{\lambda^*(\mathcal{S}, \delta, M^*(\epsilon, \delta))} \sup_{\mathbf{x} \in \mathcal{X}^n \setminus \mathcal{S}} \frac{1}{\beta_\epsilon(P^{\mathbf{x}}, P^{\mathbf{0}})}. \tag{4.22}$$

Proof. See Appendix 4.A. □

We now provide high-level explanations about the converse bounds in Lemma 36. Note that the right hand side of the upper-bounds in (4.21) and (4.22) consists of two parts. The term $\sup_{\mathbf{x} \in \mathcal{X}^n \setminus \mathcal{S}} \frac{1}{\beta_\epsilon(P^{\mathbf{x}}, P^{\mathbf{0}})}$ upper-bounds the number of codewords for any (M, ϵ) code with codewords in $\mathcal{X}^n \setminus \mathcal{S}$ [35, Theorem 31]. We use the covertness constraint to upper-bound the number of codewords in $\mathcal{S}$.

Our choice of $\mathcal{S}$ consists of the set of codewords with a specific minimum weight,

$\mathcal{S}_w \triangleq \{\mathbf{x} \in \mathcal{X}^n : \mathrm{wt}(\mathbf{x}) \geqslant w\}$. We recall here all the required quantities for $\mathcal{S}_w$. We have by [39, Eq. (191)-(193)] and [39, Lemma 11]

$$\sup_{\mathbf{x}\in\mathcal{X}^n\setminus\mathcal{S}_w} -\log\beta_\epsilon(P^{\mathbf{x}}, P^{\mathbf{0}}) \leqslant w\mathbb{D}(P_1\|P_0) - O\left(w^{\frac{1}{2}}\right), \tag{4.23}$$

$$\inf_{\alpha\in[0,1]} \left[\alpha + \beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{S}_w})\right] \leqslant 2\mathbb{Q}\left(\frac{w\sqrt{\chi_2(Q_1\|Q_0)}}{2\sqrt{n}}\right) + O\left(n^{-\frac{1}{2}}\right) + O\left(w^2 n^{-\frac{3}{2}}\right). \tag{4.24}$$

Choosing $w = \frac{2\mathbb{Q}^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\chi_2(Q_1\|Q_0)}} n^{\frac{1}{2}} - O(1)$, we obtain

$$\log \overline{M}^*(\epsilon,\delta) \leqslant Cn^{\frac{1}{2}} - \Omega\left(n^{\frac{1}{4}}\right). \tag{4.25}$$

We now prove (4.15) by first lower-bounding $\lambda^*(\mathcal{S}_w, \delta, M)$ for an appropriate choice of w.

Lemma 37. *Let $\zeta > 0$. There exists M_0 depending on δ, ζ, μ_0, and $\mathbb{D}(Q_1\|Q_0)$ such that for all $M \geqslant M_0$*

$$\lambda^*(\mathcal{S}_w, \delta, M) \geqslant \frac{1-\delta}{2} \tag{4.26}$$

for $w = \frac{\log\frac{4}{1-\delta} + \log M}{\mathbb{D}(Q_1\|Q_0) - \zeta}$.

Proof. See Appendix 4.B. □

Let $0 < \zeta < \mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0)$, M_0 be as described in Lemma 37 and $\mathcal{C}$ be an $(M^*(\epsilon,\delta), \epsilon)$ code such that $1 - \delta \geqslant \alpha(t) + \beta(t; f)$ for all t. We assume without loss of generality that $\log M^*(\epsilon,\delta) \geqslant \log M_0$, otherwise $\log M^*(\epsilon,\delta) = O(1)$. Using Lemma 37

and Lemma 36, we obtain that

$$\log M^*(\epsilon,\delta) \leqslant \log\frac{2}{1-\delta} + \sup_{\mathbf{x}\in\mathcal{X}^n\setminus\mathcal{S}_w} -\log\beta_\epsilon(P^{\mathbf{x}},P^{\mathbf{0}}) \tag{4.27}$$

where $w = \frac{\log\frac{4}{1-\delta}+\log M^*(\epsilon,\delta)}{\mathbb{D}(Q_1\|Q_0)-\zeta}$. Eq. (4.23) implies that

$$\begin{aligned}
\log M^*(\epsilon,\delta) &\leqslant \log\frac{2}{1-\delta} + w\mathbb{D}(P_1\|P_0) - O(w^{\frac{1}{2}}) && (4.28)\\
&= \log\frac{2}{1-\delta} + \frac{\log\frac{4}{1-\delta}+\log M^*(\epsilon,\delta)}{\mathbb{D}(Q_1\|Q_0)-\zeta}\mathbb{D}(P_1\|P_0)\\
&\quad - O\left(\left(\frac{\log\frac{4}{1-\delta}+\log M^*(\epsilon,\delta)}{\mathbb{D}(Q_1\|Q_0)-\zeta}\right)^{\frac{1}{2}}\right) && (4.29)\\
&= \log M^*(\epsilon,\delta)\frac{\mathbb{D}(P_1\|P_0)}{\mathbb{D}(Q_1\|Q_0)-\zeta} + O\left(\log^{\frac{1}{2}} M^*(\epsilon,\delta)\right) + O(1). && (4.30)
\end{aligned}$$

This can be re-written as

$$\left(1-\frac{\mathbb{D}(P_1\|P_0)}{\mathbb{D}(Q_1\|Q_0)-\zeta}\right)\log M^*(\epsilon,\delta) = O\left(\log^{\frac{1}{2}} M^*(\epsilon,\delta)\right). \tag{4.31}$$

By our assumption on ζ, we have $\frac{\mathbb{D}(P_1\|P_0)}{\mathbb{D}(Q_1\|Q_0)-\zeta} < 1$. Eq. (4.31) therefore leads to a contradiction if $\log M^*(\epsilon,\delta) = \omega(1)$. □

4.4.2 Covert Communication with Learning

We now provide asymptotically close lower- and upper-bounds on the number of samples ℓ required to learn a covert code.

Theorem 22. *Let* $\mathbb{D}(P_1\|P_0) < \mathbb{D}(Q_1\|Q_0)$ *and recall* $C = \frac{2\mathbb{Q}\left(\frac{1-\delta}{2}\right)\mathbb{D}(P_1\|P_0)}{\sqrt{\chi_2(Q_1\|Q_0)}}$. *If* $2\log\ell \leqslant Cn^{\frac{1}{2}} - O(n^{\frac{1}{4}})$*, we have*

$$Cn^{\frac{1}{2}} - O(n^{\frac{1}{4}}) - \log\ell \leqslant \log\overline{M}^*_\ell(\epsilon,\delta) \tag{4.32}$$

If $\log \ell = \mathbb{H}_b\left(\frac{2}{\sqrt{\chi_2(Q_1\|Q_0)}}\mathbb{Q}^{-1}\left(\frac{1-\delta}{8}\right)n^{-\frac{1}{2}}\right)n + O(1)$*, we have.*

$$\log \overline{M}^*_\ell(\epsilon,\delta) = O(1). \tag{4.33}$$

In particular, there exist two constants B_1 *and* B_2 *depending on the channels,* ϵ*, and* δ *such that if* $\log \ell \leqslant B_1 n^{\frac{1}{2}}$*, then* $\log \overline{M}^*_\ell(\epsilon,\delta) = \Omega(n^{\frac{1}{2}})$*, and if* $\log \ell \geqslant B_2 n^{\frac{1}{2}}\log n$*, then* $\log \overline{M}^*_\ell(\epsilon,\delta) = O(1)$.

Remark 9. *Note that by* (4.14) *and Theorem 21, we have* $\log \overline{M}^*_\ell(\epsilon,\delta) = Cn^{\frac{1}{2}} - \Theta\left(n^{\frac{1}{4}}\right)$ *when* $\mathbb{D}(P_1\|P_0) > \mathbb{D}(Q_1\|Q_0)$.

Achievability proof of Theorem 22. We first recall the definition of the PPM distribution from Chapter 2, which will be used in our random coding argument.

Definition 7. *Given* $w \in [\![1,n]\!]$*, let* $n = mw + r$ *for* $0 \leqslant r < w$*. We define the distribution* $P^{n,w}_{\mathbf{X},\text{PPM}}$ *on* $\mathcal{X}^n$ *as the uniform distribution on*

$$\left\{\mathbf{x}\in\mathcal{X}^n : \sum_{j=(i-1)m+1}^{im} x_j = 1\ \forall i \in [\![1,w]\!] \text{ and } \sum_{j=wm+1}^{n} x_j = 0\right\}. \tag{4.34}$$

We denote the output distribution of $P^{n,w}_{\mathbf{X},\text{PPM}}$ *for DMCs* $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ *and* $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ *by* $P^{n,w}_{\mathbf{Y},\text{PPM}}$ *and* $P^{n,w}_{\mathbf{Z},\text{PPM}}$*, respectively.*

Let $\mathbf{X}_1 \cdots, \mathbf{X}_M$ *be an iid according to* $P^{n,w}_{\mathbf{X},\text{PPM}}$*. We define a distribution* $\overline{Q}_\ell$ *over* $(\mathcal{Z}^n)^\ell$ *as*

$$\mathbb{E}_{\mathbf{X}_1\cdots,\mathbf{X}_M}\left(\left(\frac{1}{M}\sum_{i=1}^{M} Q^{\mathbf{X}_i}\right)^{\otimes\ell}\right). \tag{4.35}$$

We now state an achievabililty result in the next lemma.

Lemma 38. *Let* $\gamma, \lambda_1, \lambda_2, \lambda_3$ *be real numbers such that* $\lambda_i \geqslant 1$ *for* $i = 1,2,3$ *and* $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} < 1$*. For a test* T *with respect to* ℓ *observations,* $w = \Theta(n^{\frac{1}{2}})$*, and* M*, there exists an*

$((1-1/\lambda_3)M, \epsilon)$ *code* $\mathcal{C}$ *such that*

$$\epsilon \leqslant \mathbb{P}_{P_1^{\otimes w}}\left(\sum_{i=1}^{w} \log \frac{P_1(Y_i)}{P_0(Y_i)} \leqslant \gamma\right) + \frac{M\lambda_3\lambda_1}{\gamma} \times O(1) \tag{4.36}$$

$$\alpha_\ell(T; f) + \beta_\ell(T; f) \geqslant 1 - \frac{1}{2}\lambda_2 \left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1 - \frac{\ell+1}{\lambda_3}. \tag{4.37}$$

Proof. See Appendix 4.C. □

The spirit of the proof of Lemma 38 resembles the proof of Theorem 8 in Chapter 2, which is sketched as follows. We use a random coding argument, in which each codeword is drawn independently and from $P_{\mathbf{X},\mathrm{PPM}}^{n,w}$. As shown in Chapter 2, the probability that a message is incorrectly decoded can be upper-bounded as a sum of two terms. The first term is fixed with respect to the random coding, and the expected value of the second term is small. We further show that for a fixed test T with respect to ℓ observations, the expected value of $\alpha_\ell(T; f) + \beta_\ell(T; f)$ is $1 - \frac{1}{2}\left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1$. We use Markov inequality to conclude the existence of a code with small $\alpha_\ell(T; f) + \beta_\ell(T; f)$ and *average* probability of error. We finally expurgate the codewords with large probability of error and control how much the expurgation changes $\alpha_\ell(T; f) + \beta_\ell(T; f)$.

To use Lemma 38 to prove the achievability of Theorem 22, we need the following bounds.

Lemma 39. *We have*

$$\frac{1}{2}\left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1 \leqslant 1 - 2\mathbb{Q}\left(\frac{w}{2}\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}}\right) + \frac{\ell}{M} + O(w^{-\frac{1}{2}}). \tag{4.38}$$

For $w = 2\frac{\mathbb{Q}^{-1}\left(\frac{1-\delta}{2}\right)}{\sqrt{\chi_2(Q_1\|Q_0)}}n^{-\frac{1}{2}} - O(n^{\frac{1}{4}})$*, we have*

$$1 - 2\mathbb{Q}\left(\frac{w}{2}\sqrt{\frac{\chi_2(Q_1\|Q_0)}{n}}\right) + O(w^{-\frac{1}{2}}) \leqslant \delta - \frac{1}{\sqrt{n}} \tag{4.39}$$

For $\gamma = \mathbb{D}(P_1\|P_0)w - O(n^{\frac{1}{4}})$*, we have*

$$\mathbb{P}_{P_1^{\otimes w}}\left(\sum_{i=1}^{w}\log\frac{P_1(Y_i)}{P_0(Y_i)} \leqslant \gamma\right) \leqslant \epsilon - \frac{1}{\sqrt{n}}. \tag{4.40}$$

Proof. We prove (4.38) in Appendix 4.D. Eq. (4.39) follows from [39, Equation (250)]. Eq. (4.40) follows from [39, Lemma 5].

□

Using Lemma 38 with $\lambda_1 = n^2$, $\lambda_2 = (n+1)/n$, and $\lambda_3 = \ell n$, there exists a code $\mathcal{C}$ with probability of error bounded by

$$\epsilon - \frac{1}{\sqrt{n}} + \frac{\ell n^3 M}{\gamma} \times O(1) \tag{4.41}$$

and

$$\begin{aligned}
\alpha_\ell(T;f) + \beta_\ell(T;f) &\geqslant 1 - \frac{n+1}{2n}\left\|\overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1}\right\|_1 - \frac{\ell+1}{\ell n} && (4.42)\\
&\geqslant 1 - \left(\delta - \frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right) && (4.43)\\
&\geqslant 1 - \delta. && (4.44)
\end{aligned}$$

Finally, choosing $\log M = \log\gamma - \log\ell - 4\log n - O(1)$, we guarantee that the probability of error is less than ϵ. □

Converse proof of Theorem 22. We first provide an upper-bound on the probability of error of a generic test with respect to ℓ observations.

Lemma 40. *Let $\mathcal{H}$ be a family of functions from $\mathcal{Z}^n$ to $\{0,1\}$. Given $\mathbf{z}_1, \cdots, \mathbf{z}_\ell \in \mathcal{Z}^n$, we define*

$$t^*_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell} \triangleq \underset{h\in\mathcal{H}}{\textit{argmin}} \left[\alpha(h) + \frac{1}{\ell}\sum_{i=1}^{\ell}(1-h(\mathbf{z}_i))\right], \tag{4.45}$$

which is equivalent to a test $T(\mathbf{z}_1,\cdots,\mathbf{z}_\ell,\mathbf{z}) \triangleq t^*_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell}(\mathbf{z})$. It holds for any code $\mathcal{C}$ that

$$\alpha_\ell(T;f) + \beta_\ell(T;f) \leqslant \min_{h\in\mathcal{H}}\left[\alpha(h)+\beta(h;f)\right] + \inf_{\eta\in]0,1[}\left[2\eta + 2\left(|\mathcal{H}|+1\right)\exp\left(-2\ell\eta^2\right)\right]. \tag{4.46}$$

Proof. See Appendix 4.E. □

We now introduce a particular family of tests $\mathcal{H}$. By (4.24), there exists a test h_1 such that $Q^{\mathbf{0}}(h_1^{-1}(1)) + Q^{\mathbf{x}}(h_1^{-1}(0)) \leqslant \frac{1-\delta}{4}$ for all $\mathbf{x}$ with

$$\mathrm{wt}(\mathbf{x}) \geqslant w_1 \triangleq \frac{2}{\sqrt{\chi_2(Q_1\|Q_0)}}\mathbb{Q}^{-1}\left(\frac{1-\delta}{8}\right)n^{\frac{1}{2}} - O(1). \tag{4.47}$$

Intuitively, $h_1^{-1}(0)$ and $h_1^{-1}(1)$ are the set of typical and atypical sequences, respectively, in $\mathcal{Z}^n$ with respect to Q_0. For a non-empty $\mathcal{A}\subset\mathcal{X}^n$, we also define the test

$$\widetilde{h}_{\mathcal{A}}(\mathbf{z}) \triangleq \mathbf{1}\left\{\frac{1}{|\mathcal{A}|}\sum_{\mathbf{x}\in\mathcal{A}}\frac{Q^{\mathbf{x}}(\mathbf{z})}{Q^{\mathbf{0}}(\mathbf{z})} \geqslant \frac{8}{1-\delta}\right\} \tag{4.48}$$

Intuitively, $\widetilde{h}_{\mathcal{A}}$ performs well when the codewords are chosen from $\mathcal{A}$. We finally set $h_{\mathcal{A}}(\mathbf{z}) \triangleq \max\left(h_1(\mathbf{z}), \widetilde{h}_{\mathcal{A}}(\mathbf{z})\right)$ and $\mathcal{H} \triangleq \{h_{\mathcal{A}} : \mathcal{A}\subset\mathcal{X}^n\setminus\mathcal{S}_{w_1}, \mathcal{A}\neq\emptyset\}$,for which $|\mathcal{H}| = 2^{|\mathcal{X}^n\setminus\mathcal{S}_{w_1}|} - 1 = 2^{\sum_{i=0}^{w_1}\binom{n}{i}} - 1 \leqslant 2^{2^{\mathbb{H}_b(w_1/n)n}} - 1$ when $w_1 \leqslant n/2$. Let T be as defined in Lemma 40 for $\mathcal{H}$ and let $\mathcal{C} = (f,\phi)$ be an (M,ϵ) code with $1-\delta \leqslant \alpha(T;\mathcal{C}) + \beta(T;f)$. For w_2 specified later, we shall lower-bound the fraction of codewords with weight less than w_2, i.e.,

$$\lambda \triangleq \left|\{\mathbf{x}\in\mathrm{range}(f) : \mathrm{wt}(\mathbf{x}) < w_2\}\right|/M.$$

Lemma 40 implies that

$$1-\delta \leqslant \min_{h\in\mathcal{H}} \alpha(h)+\beta(h;f)+\inf_{\eta\in]0,1[}\left[2\eta+2\left(|\mathcal{H}|+1\right)\exp\left(-2\ell\eta^2\right)\right] \tag{4.49}$$

Let $\mathcal{A} \triangleq \{\mathbf{x}\in \mathrm{range}(f) : w_2 \leqslant \mathrm{wt}(\mathbf{x}) < w_1\} \subset \mathcal{X}^n \setminus \mathcal{S}_{w_1}$, which is such that $h_{\mathcal{A}} \in \mathcal{H}$,

$$\alpha(h_{\mathcal{A}}) = Q^{\mathbf{0}}\left(h_{\mathcal{A}}^{-1}(1)\right) \leqslant Q^{\mathbf{0}}\left(h_1^{-1}(1)\right) + Q^{\mathbf{0}}\left(\widetilde{h}_{\mathcal{A}}^{-1}(1)\right) \tag{4.50}$$

$$\leqslant Q^{\mathbf{0}}\left(h_1^{-1}(1)\right) + \frac{1-\delta}{8}, \tag{4.51}$$

and for $\zeta > 0$ and $w_2 = \frac{\log\frac{8}{1-\delta}+\log M}{\mathbb{D}(Q_1\|Q_0)-\zeta}$,

$$\beta(h_{\mathcal{A}};f) = \widehat{Q}\left(h_{\mathcal{A}}^{-1}(0)\right) \tag{4.52}$$

$$= \frac{1}{M}\sum_{m\in\mathcal{M}} Q^{f(m)}(h_{\mathcal{A}}^{-1}(0)) \tag{4.53}$$

$$\leqslant \lambda + \frac{1}{M}\sum_{m\in\mathcal{M}:w_2\leqslant \mathbf{wt}(f(m))<w_1} Q^{f(m)}(\widetilde{h}_{\mathcal{A}}^{-1}(0)) + \frac{1}{M}\sum_{m\in\mathcal{M}:w_1\leqslant \mathbf{wt}(f(m))} Q^{f(m)}(h_1^{-1}(0)) \tag{4.54}$$

$$\overset{(a)}{\leqslant} \lambda + \exp\left(-w_2\frac{\zeta^2}{\log^2\frac{1}{\mu_0}}\right) + \max_{\mathbf{x}\in\mathcal{S}_{w_1}} Q^{\mathbf{x}}(h_1^{-1}(0)), \tag{4.55}$$

where (a) follows from the same argument as in the proof of Lemma 37. When $\log M$ is large enough, we have $\exp\left(-w_2\frac{\zeta^2}{\log^2\frac{1}{\mu_0}}\right) \leqslant \frac{1-\delta}{8}$. Therefore, combining (4.51) and (4.55),

$$\alpha(h_{\mathcal{A}}) + \beta(h_{\mathcal{A}};f) \leqslant \lambda + \frac{1-\delta}{4} + Q^{\mathbf{0}}\left(h_1^{-1}(1)\right) + \max_{\mathbf{x}\in\mathcal{S}_{w_1}} Q^{\mathbf{x}}(h_1^{-1}(0)) \tag{4.56}$$

$$\overset{(a)}{\leqslant} \lambda + \frac{1-\delta}{2}, \tag{4.57}$$

where (a) follows from the definition of h_1. For $\eta = \frac{1-\delta}{16}$ and

$$\ell = \frac{\log(2) 2^{\mathbb{H}_b(w_1/n)n} + \log \frac{16}{1-\delta}}{2\left(\frac{1-\delta}{16}\right)^2}, \tag{4.58}$$

we have

$$2\left(\eta + (|\mathcal{H}| - 1)\exp\left(-2\ell\eta^2\right)\right) \leqslant \frac{1-\delta}{4} \tag{4.59}$$

Combining (4.49), (4.57), and (4.59), we obtain that $\lambda \geqslant (1-\delta)/4$. Using (4.23) and [35, Theorem 31], we have

$$\log M \leqslant \log\left(\frac{4}{1-\delta}\right) + w_2 \mathbb{D}(P_1 \| P_0) + O(w_2^{\frac{1}{2}}). \tag{4.60}$$

Repeating the argument in (4.28)-(4.30), we have $\log M = O(1)$. □

APPENDIX

4.A Proof of Lemma 36

Let us fix $\mathcal{S} \subset \mathcal{X}^n$ with $\lambda^*(\mathcal{S}, \delta) > 0$ and an (M, ϵ) code $\mathcal{C} = (f, \phi)$ with message set $\mathcal{M}$. Defining $\widetilde{\mathcal{M}} \triangleq \{m \in \mathcal{M} : f(m) \notin \mathcal{S}\} \subset \mathcal{M}$, [35, Theorem 31] implies that

$$\left|\widetilde{\mathcal{M}}\right| \leqslant \sup_{\mathbf{x} \in \mathcal{X}^n \setminus \mathcal{S}} \frac{1}{\beta_\epsilon(P^{\mathbf{x}}, P^{\mathbf{0}})} \tag{4.61}$$

To prove (4.21), it is enough to find a test $t^* : \mathcal{Z}^n \to \{0, 1\}$ independent of $\mathcal{C}$, such that $1 - \delta \leqslant \alpha(t^*) + \beta(t^*; f)$ implies that $M\overline{\lambda}^*(\mathcal{S}, \delta) \leqslant \left|\widetilde{\mathcal{M}}\right|$. By the definition of $\overline{\lambda}^*$, there exists $\alpha^* \in [0, 1]$ such that

$$\overline{\lambda}^*(\mathcal{S}, \delta) = 1 - \delta - \alpha^* - \beta^*, \tag{4.62}$$

where $\beta^* \triangleq \beta_{\alpha^*}(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x} \in \mathcal{S}})$. By the definition of $\beta_{\alpha^*}(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x} \in \mathcal{S}})$, there exists a subset $\mathcal{T}^*$ of $\mathcal{Z}^n$ such that $Q^{\mathbf{0}}(\mathcal{T}^*) \leqslant \alpha^*$ and $Q^{\mathbf{x}}(\mathcal{T}^*) \geqslant 1 - \beta^*$ for all $\mathbf{x} \in \mathcal{S}$. We define

the test $t^*(\mathbf{z}) \triangleq \mathbf{1}\{\mathbf{z} \in \mathcal{T}^*\}$ and have

$$
\begin{aligned}
1-\delta &\leqslant \beta(t^*; f) + \alpha^*(t) && (4.63)\\
&= 1 - \widehat{Q}\,(\mathcal{T}^*) + Q^{\mathbf{0}}(\mathcal{T}^*) && (4.64)\\
&\leqslant 1 - \widehat{Q}(\mathcal{T}^*) + \alpha^* && (4.65)\\
&= 1 - \frac{1}{M}\sum_{m\in\mathcal{M}} Q^{f(m)}(\mathcal{T}^*) + \alpha^* && (4.66)\\
&\overset{(a)}{\leqslant} 1 - \left(1 - \frac{\left|\widetilde{\mathcal{M}}\right|}{M}\right)(1-\beta^*) + \alpha^* && (4.67)\\
&\leqslant \alpha^* + \beta^* + \frac{\left|\widetilde{\mathcal{M}}\right|}{M} && (4.68)\\
&= 1 - \delta - \overline{\lambda}^*(\mathcal{S},\delta) + \frac{\left|\widetilde{\mathcal{M}}\right|}{M}, && (4.69)
\end{aligned}
$$

where (a) follows since $Q^{f(m)}(\mathcal{T}^*) \geqslant 0$ when $m \in \widetilde{M}$ and $Q^{f(m)}(\mathcal{T}^*) \geqslant 1-\beta^*$ otherwise. Thus, $M\lambda^*(\mathcal{S},\delta) \leqslant \left|\widetilde{\mathcal{M}}\right|$.

We now turn to the proof of (4.22). We follow the same reasoning as in the first part of the proof except that $\mathcal{T}^*$ can depend here on the code. In particular, let α^* be such that

$$\lambda^*(\mathcal{S},\delta,M) \geqslant 1-\delta-\alpha^*-\beta^*, \tag{4.70}$$

where $\beta^* \triangleq \beta_{\alpha^*}(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{S}\cap\mathrm{range}(f)})$. We then choose $\mathcal{T}^*$ such that $Q^{\mathbf{0}}(\mathcal{T}^*) \leqslant \alpha^*$ and $Q^{\mathbf{x}}(\mathcal{T}^*) \geqslant 1-\beta^*$ for all $\mathbf{x} \in \mathcal{S}\cap\mathrm{range}(f)$. The remainder of the proof is the same as the first part of the proof.

4.B Proof of Lemma 37

Let $\mathcal{A} \subset \mathcal{S}_w$ with $|\mathcal{A}| \leqslant M$. We define

$$\mathcal{T} \triangleq \left\{ \mathbf{z} \in \mathcal{Z}^n : \frac{1}{M} \sum_{\mathbf{x} \in \mathcal{A}} \frac{Q^{\mathbf{x}}(\mathbf{z})}{Q^{\mathbf{0}}(\mathbf{z})} \geqslant \frac{4}{1-\delta} \right\}. \tag{4.71}$$

Let $\widetilde{\mathbf{x}} \in \mathcal{A}$, $\widetilde{w} \triangleq \mathrm{wt}(\widetilde{\mathbf{x}})$ and $Z_1, \cdots, Z_{\widetilde{w}}$ be iid according to Q_1. We follow the ideas in the proof of [67, Lemma 3] to obtain an upper-bound on $Q^{\widetilde{\mathbf{x}}}(\mathcal{T})$ regardless of $\widetilde{\mathbf{x}}$. Note that for $\log \gamma + \log M < w\mathbb{D}(Q_1 \| Q_0)$,

$$\begin{aligned}
Q^{\widetilde{\mathbf{x}}}(\mathcal{T}) &= \mathbb{P}_{Q^{\widetilde{\mathbf{x}}}} \left(\frac{1}{M} \sum_{\mathbf{x} \in \mathcal{A}} \frac{Q^{\mathbf{x}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant \frac{4}{1-\delta} \right) && (4.72) \\
&\geqslant \mathbb{P}_{Q^{\widetilde{\mathbf{x}}}} \left(\frac{1}{M} \frac{Q^{\widetilde{\mathbf{x}}}(\mathbf{Z})}{Q^{\mathbf{0}}(\mathbf{Z})} \geqslant \frac{4}{1-\delta} \right) && (4.73) \\
&= \mathbb{P} \left(\sum_{i=1}^{\widetilde{w}} \log \frac{Q_1(Z_i)}{Q_0(Z_i)} \geqslant \log \frac{4}{1-\delta} + \log M \right) && (4.74) \\
&\overset{(a)}{\geqslant} 1 - \exp \left(- \frac{\left(\widetilde{w}\mathbb{D}(Q_1 \| Q_0) - \log \frac{4}{1-\delta} - \log M \right)^2}{\widetilde{w} \log^2 \frac{1}{\mu_0}} \right), && (4.75)
\end{aligned}$$

where (a) follows from Hoeffding's inequality (Theorem 3 in Chapter 1). Since $\widetilde{w} \geqslant w = \frac{\log \frac{4}{1-\delta} + \log M}{\mathbb{D}(Q_1 \| Q_0) - \zeta}$, we have

$$\begin{aligned}
& \exp \left(- \frac{\left(\widetilde{w}\mathbb{D}(Q_1 \| Q_0) - \log \frac{4}{1-\delta} - \log M \right)^2}{\widetilde{w} \log^2 \frac{1}{\mu_0}} \right) && (4.76) \\
& \leqslant \exp \left(- \frac{\left(\widetilde{w}\mathbb{D}(Q_1 \| Q_0) - \widetilde{w}(\mathbb{D}(Q_1 \| Q_0) - \zeta) \right)^2}{\widetilde{w} \log^2 \frac{1}{\mu_0}} \right) && (4.77) \\
& = \exp \left(- \frac{\widetilde{w}\zeta^2}{\log^2 \frac{1}{\mu_0}} \right) && (4.78) \\
& \leqslant \exp \left(- \frac{w\zeta^2}{\log^2 \frac{1}{\mu_0}} \right). && (4.79)
\end{aligned}$$

Additionally, by the definition of $\mathcal{T}$,

$$Q^{\mathbf{0}}(\mathcal{T}) \leqslant \frac{1-\delta}{4M} \sum_{\mathbf{x}\in\mathcal{A}} Q^{\mathbf{x}}(\mathcal{T}) \leqslant \frac{(1-\delta)|\mathcal{A}|}{4M} \leqslant \frac{1-\delta}{4}. \tag{4.80}$$

By the definition of $\beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{A}})$, we have

$$\beta_{\frac{1-\delta}{4}}(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{A}}) \leqslant \exp\left(-w\frac{\zeta^2}{\log^2 \frac{1}{\mu_0}}\right), \tag{4.81}$$

which is less than $\frac{1-\delta}{4}$ for M large enough. Therefore,

$$\inf_{\alpha\in[0,1]} \left[\alpha + \beta_\alpha(Q^{\mathbf{0}}, \{Q^{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{A}})\right] \leqslant \frac{1-\delta}{2}. \tag{4.82}$$

4.C Proof of Lemma 38

We first prove a technical lemma.

Lemma 41. *Let f_1 and f_2 be two encoders inducing output distributions $\widehat{Q}_1$ and $\widehat{Q}_2$, respectively. We have*

$$|\alpha_\ell(T; f_1) + \beta_\ell(T; f_1) - \alpha_\ell(T; f_2) - \beta_\ell(T; f_2)| \leqslant \ell \left\| \widehat{Q}_1 - \widehat{Q}_2 \right\|_1 \tag{4.83}$$

Proof. We first show that $|\alpha_\ell(T; f_1) - \alpha_\ell(T; f_2)| \leqslant \ell/M$. Note that

$$\begin{aligned} |\alpha_\ell(T; f_1) - \alpha_\ell(T; f_2)| &= |(\widehat{Q}_1^{\otimes\ell} \otimes Q^{\mathbf{0}})(T^{-1}(1)) - (\widehat{Q}_2^{\otimes\ell} \otimes Q^{\mathbf{0}})(T^{-1}(1))| && (4.84)\\ &\leqslant \frac{1}{2}\left\| \widehat{Q}_1^{\otimes\ell} \otimes Q^{\mathbf{0}} - \widehat{Q}_2^{\otimes\ell} \otimes Q^{\mathbf{0}} \right\|_1 && (4.85)\\ &\leqslant \frac{\ell}{2}\left\| \widehat{Q}_1 - \widehat{Q}_2 \right\|_1. && (4.86) \end{aligned}$$

We can similarly show that $|\beta_\ell(T; f_1) - \beta_\ell(T; f_2)| \leqslant \frac{1}{2}(\ell+1)\left\| \widehat{Q}_1 - \widehat{Q}_2 \right\|$; using the triangle inequality completes the proof. □

Let $\mathcal{M} \triangleq [\![1, M]\!]$ and $F : \mathcal{M} \to \mathcal{X}^n$ be a random encoder such that each codeword is distributed according to $P^{n,w}_{\mathbf{X},\text{PPM}}$. We define $\widehat{Q} \triangleq \frac{1}{M}\sum_{m\in\mathcal{M}} Q^{F(m)}$, a *random* PMF over $\mathcal{Z}^n$. We now explain the decoding rule, which is based on [13]. Given $\gamma > 0$, upon observing $\mathbf{y}$, the decoder forms the estimate $\widehat{W} = m$ of the transmitted message W if there exists a unique $m \in \mathcal{M}$ such that

$$\log \frac{P^{F(m)}(\mathbf{y})}{P^{\mathbf{0}}(\mathbf{y})} > \gamma. \tag{4.87}$$

If there is no w satisfying (4.87), the decoder declares an error. The conditional probability of error when $W = m$ is upper-bounded [35, Eq. (72)] by $\epsilon^{(1)}_m + \epsilon^{(2)}_m$ where

$$\epsilon^{(1)}_m \triangleq \sum_{\mathbf{y}} P^{F(m)}(\mathbf{y})\mathbf{1}\{\log \frac{P^{F(m)}(\mathbf{y})}{P^{\mathbf{0}}(\mathbf{y})} \leqslant \gamma\} \tag{4.88}$$

$$\epsilon^{(2)}_m \triangleq \sum_{m'\in\mathcal{M}:m'\neq m} \sum_{\mathbf{y}} P^{F(m)}(\mathbf{y})\mathbf{1}\{\log \frac{P^{F(m')}(\mathbf{y})}{P^{\mathbf{0}}(\mathbf{y})} \geqslant \gamma\}. \tag{4.89}$$

As discussed in [39], $\epsilon^{(1)}_m = \mathbb{P}_{P_1^{\otimes w}}\left(\sum_{i=1}^w \log \frac{P_1(Y_i)}{P_0(Y_i)} \leqslant \gamma\right)$ irrespective of F because of our specific choice of input distribution $P^{n,w}_{\mathbf{X},\text{PPM}}$. We moreover have by [39, Eq. (42) and (133)],

$$\mathbb{E}_F\left(\epsilon^{(2)}_m\right) \leqslant \frac{M}{\gamma}\mathbb{E}_{P^{n,w}_{\mathbf{Y},\text{PPM}}}\left(\frac{P^{n,w}_{\mathbf{Y},\text{PPM}}(\mathbf{Y})}{P^{\mathbf{0}}(\mathbf{Y})}\right) \tag{4.90}$$

$$\leqslant \frac{M}{\gamma}\exp\left(\frac{w(w+1)}{n}\chi_2(P_1\|P_0)\right). \tag{4.91}$$

We therefore have for $\lambda_1 > 1$,

$$\mathbb{P}_F\left(\frac{1}{M}\sum_{m\in\mathcal{M}} \epsilon^{(2)}_m \geqslant \lambda_1\frac{M}{\gamma}\exp\left(\frac{w(w+1)}{n}\chi_2(P_1\|P_0)\right)\right) \leqslant \frac{1}{\lambda_1}. \tag{4.92}$$

We further have

$$\mathbb{E}_F(\alpha(T;F)+\beta(T;F)) = \mathbb{E}_F\Big((\widehat{Q}^{\otimes \ell}\otimes Q^{\mathbf{0}})(T^{-1}(0)) + (\widehat{Q}^{\otimes \ell+1})(T^{-1}(1))\Big) \tag{4.93}$$

$$= (\overline{Q}_\ell \otimes Q^{\mathbf{0}})(T^{-1}(0)) + \overline{Q}_{\ell+1}(T^{-1}(1)) \tag{4.94}$$

$$\overset{(a)}{\geqslant} 1 - \frac{1}{2}\big\|\overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1}\big\|_1 \tag{4.95}$$

where (a) follows from $\frac{1}{2}\big\|\overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1}\big\|_1 = \sup_{\mathcal{T}}\left[(\overline{Q}_\ell \otimes Q^{\mathbf{0}})(\mathcal{T}) - \overline{Q}_{\ell+1}(\mathcal{T})\right]$ Using Markov inequality implies that for $\lambda_2 > 1$

$$\mathbb{P}_F\left(\alpha(T;F)+\beta(T;F) \leqslant 1 - \frac{1}{2}\lambda_2\big\|\overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1}\big\|_1\right) \leqslant \frac{1}{\lambda_2} \tag{4.96}$$

Thus, there exists a code $\mathcal{C}$ with message set $\mathcal{M}$ such that

$$\epsilon_m^{(1)} = \mathbb{P}_{P_1^{\otimes w}}\left(\sum_{i=1}^{w}\log\frac{P_1(Y_i)}{P_0(Y_i)} \leqslant \gamma\right), \forall m \in \mathcal{M} \tag{4.97}$$

$$\frac{1}{M}\sum_{m\in\mathcal{M}}\epsilon_m^{(2)} \leqslant \lambda_1\frac{M}{\gamma}\exp\left(\frac{w(w+1)}{n}\chi_2(P_1\|P_0)\right) \tag{4.98}$$

$$\alpha_\ell(T;f)+\beta_\ell(T;f) \geqslant 1 - \frac{1}{2}\lambda_2\big\|\overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1}\big\|_1 \tag{4.99}$$

for all λ_1, λ_2 with $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} < 1$. We now expurgate all messages with

$$\epsilon_m^{(2)} \geqslant \lambda_3\lambda_1\frac{M}{\gamma}\exp\left(\frac{w(w+1)}{n}\chi_2(P_1\|P_0)\right) \tag{4.100}$$

to obtain a new code $\widetilde{\mathcal{C}} = (\widetilde{f}, \widetilde{\phi})$ with message set $\widetilde{\mathcal{M}}$. Note that

$$\max_{m \in \widetilde{\mathcal{M}}} \epsilon_m^{(1)} + \epsilon_m^{(2)} \leqslant \mathbb{P}_{P_1^{\otimes w}} \left(\sum_{i=1}^{w} \log \frac{P_1(Y_i)}{P_0(Y_i)} \leqslant \gamma \right) + \frac{\lambda_3 \lambda_1 M}{\gamma} \exp\left(\frac{w(w+1)}{n} \chi_2(P_1 \| P_0) \right) \tag{4.101}$$

$$\left| \widetilde{\mathcal{M}} \right| \geqslant \left(1 - \frac{1}{\lambda_3} \right) M. \tag{4.102}$$

Let $\widehat{Q}_1$ and $\widehat{Q}_2$ denote the induced output distribution of $\mathcal{C}$ and $\widetilde{C}$, respectively. We have $\widehat{Q}_2 = (1 - 1/\lambda_3)\widehat{Q}_1 + \lambda_3 Q'$ for a PMF Q' over $\mathcal{Z}^n$. Hence, $\frac{1}{2}\left\| \widehat{Q}_1 - \widehat{Q}_2 \right\|_1 \leqslant \frac{1}{2\lambda_3}$. Using Lemma 41 together with (4.99), we obtain that

$$\alpha_\ell(T; \widetilde{f}) + \beta_\ell(T; \widetilde{f}) \geqslant 1 - \frac{1}{2}\lambda_2 \left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1 - \frac{\ell+1}{\lambda_3}. \tag{4.103}$$

4.D Proof of Lemma 39

Let $F : [\![1, M]\!] \to \mathcal{X}^n$ be a random function such that $F(1), \cdots, F(M)$ be iid according to $P_{\mathbf{X},\mathrm{PPM}}^{n,w}$ and $\widehat{Q} \triangleq \frac{1}{M}\sum_{m=1}^{M} Q^{F(m)}$ be a random PMF over $\mathcal{Z}$. We have $\overline{Q}_\ell = \mathbb{E}_F\left(\widehat{Q}^{\otimes \ell}\right)$ by definition. Therefore,

$$\left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1 \tag{4.104}$$

$$= \sum_{\mathbf{z}_1, \cdots, \mathbf{z}_\ell, \mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} \widehat{Q}(\mathbf{z}_i) \left(\widehat{Q}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right) \right| \tag{4.105}$$

$$\leqslant \frac{1}{M^{\ell+1}} \sum_{m_1, \cdots, m_\ell, m} \sum_{\mathbf{z}_1, \cdots, \mathbf{z}_\ell, \mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \left(Q^{F(m)}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right) \right|. \tag{4.106}$$

Suppose that $m \neq m_i$ for all $i \in [\![1, \ell]\!]$. The terms $\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i)$ and $Q^{F(m)}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z})$ are then independent, which results in

$$\sum_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell,\mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \left(Q^{F(m)}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right) \right| \tag{4.107}$$

$$= \sum_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell,\mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \right) \mathbb{E}_F \left(Q^{F(m)}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right| \tag{4.108}$$

$$= \sum_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell,\mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \right) \left(P^{n,w}_{\mathbf{Z},\text{PPM}}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right| \tag{4.109}$$

$$= \sum_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell} \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \right) \sum_{\mathbf{z}} \left| \left(P^{n,w}_{\mathbf{Z},\text{PPM}}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right| \tag{4.110}$$

$$= \left\| P^{n,w}_{\mathbf{Z},\text{PPM}} - Q^{\mathbf{0}} \right\|_1. \tag{4.111}$$

We also have for all $m_1, \cdots, m_\ell, m$,

$$\sum_{\mathbf{z}_1,\cdots,\mathbf{z}_\ell,\mathbf{z}} \left| \mathbb{E}_F \left(\prod_{i=1}^{\ell} Q^{F(m_i)}(\mathbf{z}_i) \left(Q^{F(m)}(\mathbf{z}) - Q^{\mathbf{0}}(\mathbf{z}) \right) \right) \right| \leqslant 2. \tag{4.112}$$

Therefore,

$$\frac{1}{2} \left\| \overline{Q}_\ell \otimes Q^{\mathbf{0}} - \overline{Q}_{\ell+1} \right\|_1 \leqslant \frac{1}{2} \left\| P^{n,w}_{\mathbf{Z},\text{PPM}} - Q^{\mathbf{0}} \right\|_1 + \frac{\sum_{m_1,\cdots,m_\ell,m} \mathbf{1}\{\exists i \in [\![1, \ell]\!] : m = m_i\}}{M^{\ell+1}} \tag{4.113}$$

$$\leqslant \frac{1}{2} \left\| P^{n,w}_{\mathbf{Z},\text{PPM}}, Q^{\mathbf{0}} \right\|_1 + \frac{\ell}{M}. \tag{4.114}$$

Finally, we have $\frac{1}{2} \left\| P^{n,w}_{\mathbf{Z},\text{PPM}}, Q^{\mathbf{0}} \right\|_1 \leqslant 1 - 2\mathbb{Q}\left(\frac{w}{2} \sqrt{\frac{\chi_2(Q_1 \| Q_0)}{n}} \right) + O(w^{-\frac{1}{2}})$ by [39, Lemma 8].

4.E Proof of Lemma 40

Let $\mathbf{Z}_1, \cdots, \mathbf{Z}_\ell$ be iid according to $\widehat{Q}$. Let $h^* \triangleq \operatorname{argmin}_{h\in\mathcal{H}} [\alpha(h) + \beta(h; f)]$ and define two events

$$\mathcal{E}_1 \triangleq \{\beta(h^*; f) \geqslant \frac{1}{\ell}\sum_{i=1}^{\ell}(1 - h^*(\mathbf{z}_i)) - \eta\}, \tag{4.115}$$

$$\mathcal{E}_2 \triangleq \{\beta(h; f) \leqslant \frac{1}{\ell}\sum_{i=1}^{\ell}(1 - h(\mathbf{z}_i)) + \eta, \forall h \in \mathcal{H}\}. \tag{4.116}$$

By the law of total probability, we have

$$\alpha_\ell(T; f) + \beta_\ell(T; f) = \mathbb{E}\big(\alpha(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}) + \beta(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}; f)\big) \tag{4.117}$$

$$\leqslant \mathbb{E}\big(\alpha(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell})\big) + \mathbb{E}\big(\beta(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}; f)|\mathcal{E}_1 \cap \mathcal{E}_2\big) + 2\mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c). \tag{4.118}$$

We upper-bound the first term as

$$\mathbb{E}\big(\alpha(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}) + \beta(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}; f)|\mathcal{E}_1 \cap \mathcal{E}_2\big) \tag{4.119}$$

$$\overset{(a)}{\leqslant} \mathbb{E}\left(\alpha(t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}) + \frac{1}{\ell}\sum_{i=1}^{\ell}(1 - t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}(\mathbf{Z}_i))|\mathcal{E}_1 \cap \mathcal{E}_2\right) + \eta \tag{4.120}$$

$$\overset{(b)}{\leqslant} \mathbb{E}\left(\alpha(h^*) + \frac{1}{\ell}\sum_{i=1}^{\ell}(1 - h^*(\mathbf{Z}_i))|\mathcal{E}_1 \cap \mathcal{E}_2\right) + \eta \tag{4.121}$$

$$\overset{(c)}{\leqslant} \alpha(h^*) + \beta(h^*; f) + 2\eta \tag{4.122}$$

$$= \operatorname*{argmin}_{h\in\mathcal{H}} [\alpha(h) + \beta(h; f)] + 2\eta, \tag{4.123}$$

where (a) follows from the definition of $\mathcal{E}_2$, (b) follows from the definition of $t^*_{\mathbf{Z}_1,\cdots,\mathbf{Z}_\ell}$, and (c) follows from the definition of $\mathcal{E}_1$. Finally the Hoeffding inequality (Theorem 3 in

Chapter 1) and the union bound yield that

$$\mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) \leqslant (|\mathcal{H}| + 1) \exp\left(-2\ell\eta^2\right). \tag{4.124}$$

CHAPTER 5

COVERT SECRET KEY GENERATION: PASSIVE AND ACTIVE WARDENS

5.1 Summary

We investigate the problem of covert and secret key generation over a state-dependent discrete memoryless channel with one-way public discussion in which an adversary, the warden, may arbitrarily choose the channel state. We develop an adaptive protocol that, under conditions that we explicitly specify, not only allows the transmitter and the legitimate receiver to exchange a secret key but also conceals from the active warden whether the protocol is being run. When specialized to passive adversaries that do not control the channel state, we partially characterize the covert secret key capacity. In particular, the covert secret key capacity is sometimes equal to the covert capacity of the channel, so that secrecy comes "for free." The content of this chapter is based on [68, 69].

5.2 Introduction

Following the early results of Ahlswede and Csiszár [70] and Maurer [71], secret key generation from correlated observations using an authenticated public channel has attracted significant attention, especially in the context of wireless channels [72, 73]. We investigate here the problem of *covert* secret key generation, in which legitimate parties must not only agree on a common secret key but also keep the key generation protocol undetectable by an adversary, referred to as the *warden*. Our work partially addresses the question by showing that, under conditions that we shall precisely specify, covert and secret key generation is possible even with an active warden. Specifically, the results reported here extend our preliminary results restricted to a passive warden [69] to a model with an active warden who can arbitrarily vary the channel state, except when no information is sent on the main

channel. While this restriction arises from the technicalities in our proofs, it is justified in certain practical scenarios. Specifically, for wireless channels in which the action of the warden corresponds to tampering with the gain of the legitimate receiver, the gain has no effect when no signal is transmitted because fading acts as a multiplicative coefficient.

As in most results on secret key generation, the presence of an authenticated public communication is pivotal in our coding scheme to enable covert secret key generation and therefore covert communication for channels over which a secret key is required [13]. To avoid improving the warden's detection ability, we impose a probability distribution on the public communication and make certain that the warden cannot detect the communication with any test jointly performed on the observations of the noisy channel and of the public channel. This model relates to *stealth secret key generation* from a source model [74]; however, stealth is a less stringent requirement than covertness, so that our results are of a different nature and exploit different proof techniques to characterize the covert secret key capacity. We emphasize that our approach differs from previous studies on two accounts. First, we neither impose any limit on the warden's actions nor consider any statistical model for the channel states, so that the warden may take any action and our coding scheme remains reliable and covert for all possible state sequences. Second, the covert throughput is adapted to the warden's actions, i.e., legitimate parties decide how many bits are extracted based on the quality of the channels, and we use ideas for estimation that we introduced in [75] in the context of learning over wiretap channels.

The remainder of the chapter is organized as follows. In Section 5.3, we formally introduce our model for covert secret key generation. In Section 5.4 and Section 5.5, we develop our results on covert secret key generation for passive and active models, respectively.

5.3 Problem Formulation

We consider the channel model for secret key generation illustrated in Fig. 5.3.1, in which two legitimate parties, Alice and Bob, attempt to generate a secret key while keeping the

entire key generation process undetectable by a warden Willie. The channel is a state-dependent DMC $(\mathcal{X} \times \mathcal{S}, W_{YZ|XS}, \mathcal{Y} \times \mathcal{Z})$, in which the state S is under Willie's control while the input X is under Alice's control. Bob and Willie's channel outputs are Y and Z, respectively. For simplicity, we assume that $\mathcal{X} = \mathcal{S} \triangleq \{0, 1\}$, where 0 is the input corresponding to the absence of communication. For $x, s \in \{0, 1\}$, we define

$$P_x^s \triangleq W_{Y|X=x,S=s}, \quad Q_x^s \triangleq W_{Z|X=x,S=s}, \tag{5.1}$$

$$\text{and } (PQ)_x^s \triangleq W_{YZ|X=x,S=s}. \tag{5.2}$$

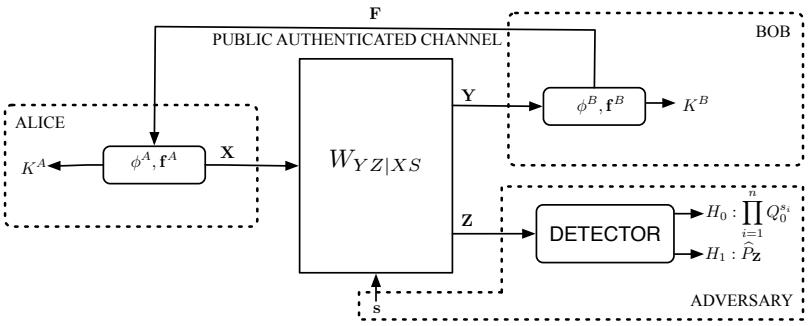

Figure 5.3.1: Covert secret key generation model

Secret key generation is enabled by the presence of a public authenticated link of unlimited capacity, which Bob may use to transmit symbols in alphabet $\mathcal{F}$. Alice and Bob have access to a source of secret common randomness $(\mathcal{R}_C, P_{R_C})$ and each possess a local source of randomness denoted by $(\mathcal{R}_A, P_{R_A})$, $(\mathcal{R}_B, P_{R_B})$, respectively. To make the problem non-trivial, the entropy of P_{R_C} is subtracted from the number of generated key bits in our throughput analysis.

In the presence of an active attacker controlling the channel state, the final length of the key is not known ahead of time. Alice and Bob must therefore merely agree a priori on a maximum number of channel uses and bits of secret key. Formally, a code for key generation $\mathcal{C}$ over n channel uses for a maximum of m key bits consists of the following.

- n encoding function for Alice $\mathbf{f}^A = (f_1^A, \cdots, f_n^A)$, where $f_i^A : \mathcal{F}^{i-1} \times \mathcal{R}^A \times \mathcal{R}^C \to \mathcal{X}$

specifies the symbol sent by Alice at time $i \in [\![1, n]\!]$;

- n encoding functions for Bob $\mathbf{f}^B = (f_1^B, \cdots, f_n^B)$, where $f_i^B : \mathcal{Y}^i \times \mathcal{R}^B \times \mathcal{R}^C \to \mathcal{F}$ specifies the symbol sent by Bob at time $i \in [\![1, n]\!]$ over the public channel;
- a key extraction function $\phi^A : \mathcal{F}^n \times \mathcal{R}^A \times \mathcal{R}^C \to \{0,1\}^m$ for Alice;
- a key extraction function $\phi^B : \mathcal{Y}^n \times \mathcal{R}^B \times \mathcal{R}^C \to \{0,1\}^m$ for Bob;
- an estimator for the number of key bits $\ell^A : \mathcal{F}^n \times \mathcal{R}^A \times \mathcal{R}^C \to [\![0, m]\!]$ for Alice;
- an estimator for the number of key bits $\ell^B : \mathcal{Y}^n \times \mathcal{R}^B \times \mathcal{R}^C \to [\![0, m]\!]$ for Bob.

We assume that the protocol is known to all parties. The sequence of n random symbols transmitted by Alice is denoted $\mathbf{X} \in \mathcal{X}^n$, while the sequence of states is denoted $\mathbf{s} \in \mathcal{S}^n$. Note that we do not require the existence of a probability distribution for the state sequence. The sequence of observations at Bob and Willie are denoted $\mathbf{Y} \in \mathcal{Y}^n$ and $\mathbf{Z} \in \mathcal{Z}^n$, respectively. Bob's public communication is collectively denoted by $\mathbf{F}$ and the generated keys are denoted K_A, K_B, respectively. For a fixed state sequence $\mathbf{s}$, the distribution induced by the coding scheme is denoted $\widehat{P}_{\mathbf{XYZF}K^AK^B\ell^A\ell^B|\mathbf{s}}$. Note that our model allows for an arbitrary $\mathbf{s}$ but does not allow the warden to adapt $\mathbf{s}$ to its observations.

The performance of the key generation scheme is measured in terms of the following metrics:

- the probability of error

$$P_e(\mathcal{C}|\mathbf{s}) \triangleq \mathbb{P}_{\widehat{P}_{K^AK^B}}\big(\ell^A \neq \ell^B \text{ or } K^A(i) \neq K^B(i) \text{ for } i \in [\![1, \ell^B]\!]\big|\mathbf{s}\big), \tag{5.3}$$

  where $K^A(i)$ and $K^B(i)$ denote the i^{th} bit in Alice's key and Bob's key, respectively;

- the secrecy $\mathbb{D}\Big(\widehat{P}_{K^B\mathbf{FZ}|\mathbf{s}} \| P_{K^B}^{\text{unif}} \otimes \widehat{P}_{\mathbf{FZ}|\mathbf{s}}\Big)$;
- the covertness $C(\mathcal{C}|\mathbf{s}) \triangleq \mathbb{D}\Big(\widehat{P}_{\mathbf{FZ}|\mathbf{s}} \| P_{\mathbf{F}}^{\text{unif}} \otimes \prod_{i=1}^n Q_0^{s_i}\Big)$.

We call $\mathcal{C}$ a $(2^m, n, r, \epsilon, \delta, \tau, \mathbf{s})$ code if $\mathbb{H}\big(R^C\big) \leqslant r$, $P_e(\mathcal{C}|\mathbf{s}) \leqslant \epsilon$, $S(\mathcal{C}|\mathbf{s}) \leqslant \delta$, and $C(\mathcal{C}|\mathbf{s}) \leqslant \tau$.

While the definitions of the probability of error and secrecy metrics are standard, there are somewhat arbitrary choices in our definition of covertness. We require Alice's transmission to be indistinguishable from an all-0 transmission, but we allow Bob to send symbols on the public channel as long as their observation does not help Willie's detection. For small $\tau > 0$, the covertness constraint ensures that the public communication is nearly uniformly distributed and independent of the Willie's observation on the noisy channel. One can think of Bob as a terminal emitting seemingly random "beacons" that do not divulge the existence of a secret key generation protocol.

The throughput achieved by a protocol is defined as follows.

Definition 8. *A throughput R is achievable with respect to the sequence $\{\mathbf{s}_n\}_{n\geqslant 1}$, if there exists a sequence of $(2^{m_n}, n, r_n, \epsilon_n, \delta_n, \tau_n, \mathbf{s}_n)$ codes $\{\mathcal{C}_n\}_{n\geqslant 1}$ such that*

$$\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \tau_n = 0, \; m_n = \omega(\log n) \tag{5.4}$$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\ell_n^B - r_n}{\sqrt{n\tau_n}} \geqslant R \right) = 1. \tag{5.5}$$

The special case of a *passive attacker* consists of the situation in which the state sequence is fixed and known ahead of time. We set this sequence to be $\mathbf{s} = \mathbf{0}$ and drop the indices referring to the state to simplify notation. In this case, note that the estimators ℓ^A and ℓ^B are not needed and that the total number of key bits m may be fixed ahead of time. We can then formally define the covert secret key capacity as follows.

Definition 9. *A throughput R is achievable with a passive attacker if there exists a sequence*

of $(2^{m_n}, n, 0, \epsilon_n, \delta_n, \tau_n)$ *codes* $\{\mathcal{C}_n\}_{n\geqslant 1}$ *such that*

$$\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \tau_n = 0,\ m_n = \omega(\log n), \tag{5.6}$$

$$\liminf_{n\to\infty} \frac{m_n}{\sqrt{n\tau_n}} \geqslant R. \tag{5.7}$$

The supremum of all achievable throughputs is denoted C_{csk}.

Remark 10. *Our restriction to* $|\mathcal{S}| = |\mathcal{X}| = 2$ *simplifies the technical details in our proofs. By following [13, Section VII-B], one can extend the results to any finite* $\mathcal{X}$. *When* $|S| > 2$, *one can adapt our estimation protocol to operate on the type of the state sequences instead of their weight.*

5.4 Covert Secret Key Capacity with a Passive Warden

For completeness, we recall without proof the partial characterization of the covert secret key capacity in the presence of a passive warden, which is our main result from [69].

Theorem 23. *If* $(PQ)_0 = P_0 \otimes Q_0$, *then*

$$\begin{aligned}\sqrt{\frac{2}{\chi_2(Q_1\|Q_0)}}\left(\mathbb{D}((PQ)_1\|(PQ)_0) - \mathbb{D}(Q_1\|Q_0)\right) &\geqslant C_{\text{csk}} \\ \geqslant \sqrt{\frac{2}{\chi_2(Q_1\|Q_0)}}\left(\mathbb{D}((PQ)_1\|(PQ)_0) - \mathbb{D}(Q_1\|Q_0) - \mathbb{D}((PQ)_1\|P_1\otimes Q_1)\right).\end{aligned} \tag{5.8}$$

Proof. See [69]. □

As an application of the above result, we characterize the exact covert secret key capacity when the channels from Alice to Bob and Willie are independent.

Corollary 1. *If* $(PQ)_1 = P_1 \otimes Q_1$ *and* $(PQ)_0 = P_0 \otimes Q_0$*, then*

$$C_{\text{csk}} = \sqrt{\frac{2}{\chi_2(Q_1 \| Q_0)}} \mathbb{D}(P_1 \| P_0). \tag{5.9}$$

Corollary 1 may be somewhat surprising in that it suggests that secrecy comes "for free" since the covert secret-key capacity is equal to the covert capacity of the channel. In practice, however, some small amount of privacy amplification would still be needed and the effect of the warden's channel only disappears in the asymptotic limit of large sequences. This result is an artifact of the model, which ensures that the information leakage from Bob to Willie has *negligible* scaling compared to the information transfer from Bob to Alice.

5.5 Covert Throughput with an Active Warden

We now develop results for an active warden and show the existence of a sequence of coding schemes generating a key for *any* sequence of states. The number of generated key bits depends on the state sequences only through their weights.

Theorem 24. *Let* $(\mathcal{X} \times \mathcal{S}, W_{YZ|XS}, \mathcal{Y}, \mathcal{Z})$ *be an arbitrarily varying DMC with* $\mathcal{X} = \mathcal{S} = \{0, 1\}$*,* $P_0^0 = P_0^1 = P_0$*,* $P_1^1 \neq P_1^0$*, and* $(PQ)_0^s = P_0^s \otimes Q_0^s$ *for* $s \in \mathcal{S}$*. There exists a sequence of codes* $\{\mathcal{C}_n\}_{n \geqslant 1}$ *such that for all* $\beta \in [0, 1]$ *and all sequences* $\{\mathbf{s}_n\}_{n \geqslant 1}$ *with* $\lim_{n \to \infty} \frac{wt(\mathbf{s}_n)}{n} = \beta$*, the following covert throughput is achievable*

$$R(\beta) \triangleq \sqrt{2} \frac{\mathbb{D}((1-\beta)P_1^0 + \beta P_1^1 \| P_0) - ((1-\beta)I^0 + \beta I^1)}{\sqrt{(1-\beta)\chi_2(Q_1^0 \| Q_0^0) + \beta \chi_2(Q_1^1 \| Q_0^1)}}, \tag{5.10}$$

where for $s \in \mathcal{S}$*,*

$$I^s \triangleq \mathbb{D}(Q_1^s \| Q_0^s) + \mathbb{D}(P_1^s \| P_0^s) - \mathbb{D}((PQ)_1^s \| (PQ)_0^s) + \mathbb{D}((PQ)_1^s \| P_1^s \otimes Q_1^s). \tag{5.11}$$

Remark 11. *Comparing the throughputs in* (5.8) *and* (5.10)*, note that the quantities cor-*

responding to the main channel in (5.8) *are replaced by the same quantities for the channel* $\sum_{i=1}^{n} \frac{1}{n} W_{Y|XS=s_i}$*, and the quantities corresponding to the warden's channel are replaced by the average of those quantities over the different channel uses. The intuition is that Alice and Bob have no direct access to the state sequences and only approximate the weight through their noisy observations. From their perspective, the best approximation of the main channel is* $\sum_{i=1}^{n} \frac{1}{n} W_{Y|XS=s_i}$*. In contrast, Willie knows the exact state sequence therefore obtains information from each realized channel use.*

Remark 12. *For an AVC, one can eliminate the common randomness when the channel is non-symmetrizable as the capacity without common randomness is non-zero [76]. However, such consideration does not directly apply to our model, which includes a feedback link from the receiver.*

We also establish a straightforward converse result for Theorem 24. Note that it is challenging to obtain a converse that matches our achievability result for a couple of reasons. First, even without covertness constraint and active adversaries, it is generally hard to provide a tight converse for the secret key generation problem because of the interactions allowed by the existence of the public communication. Second, our achievability result proves the possibility of covert secret key generation for all channel state sequences, which makes the problem of finding a tight converse more difficult.

Theorem 25. *For any sequence of codes* $\{\mathcal{C}_n\}_{n\geqslant 1}$ *that achieves covert throughput* R *over the channel in Theorem 24 for a state sequence* $\{\mathbf{s}_n\}_{n\geqslant 1}$ *with* $\lim_{n\to\infty} \frac{wt(\mathbf{s}_n)}{n} = \beta$*, we have*

$$R \leqslant \sqrt{2}\sqrt{\beta \frac{(\mathbb{D}(P_1^1\|P_0) - I^1)^2}{\chi_2(Q_1^0\|Q_0^0)} + (1-\beta)\frac{(\mathbb{D}(P_1^0\|P_0) - I^0)^2}{\chi_2(Q_1^1\|Q_0^1)}}, \tag{5.12}$$

where I^0 *and* I^1 *are defined in Theorem 24.*

Proof. To obtain a converse, we can assume that Alice and Bob know the state sequence $\mathbf{s}_n = (s_1, \cdots, s_n)$ so that the result follows from the same argument as for the passive

warden [69]. In particular, let $\mathcal{C}_n$ be a $(2^m, n, r, \epsilon, \delta, \tau, \mathbf{s})$. By [70], we obtain that

$$\sqrt{n\tau_n}R \leqslant (1+o(1))\sum_{i=1}^{n}\mathbb{I}(X_i;Y_i|Z_iS_i=s_i). \tag{5.13}$$

Let J be uniformly distributed over $[\![1,n]\!]$ and $\mu^s = \mathbb{P}(X_J=1|S_J=s)$. Using [69, Eq. (60) and Eq. (61)] and $\mathrm{wt}(\mathbf{s}_n) = n(\beta+o(1))$, we obtain

$$R \leqslant (1+o(1))\frac{\beta\mu^1\left(\mathbb{D}(P_1^1\|P_0)-I^1\right)+(1-\beta)\mu^0\left(\mathbb{D}(P_1^0\|P_0)-I^0\right)}{\sqrt{\beta(\mu^1)^2\chi_2(Q_1^1\|Q_0^1)+(1-\beta)(\mu^0)^2\chi_2(Q_1^0\|Q_0^0)}}. \tag{5.14}$$

Maximizing over μ^1 and μ^0 yields the desired result.

□

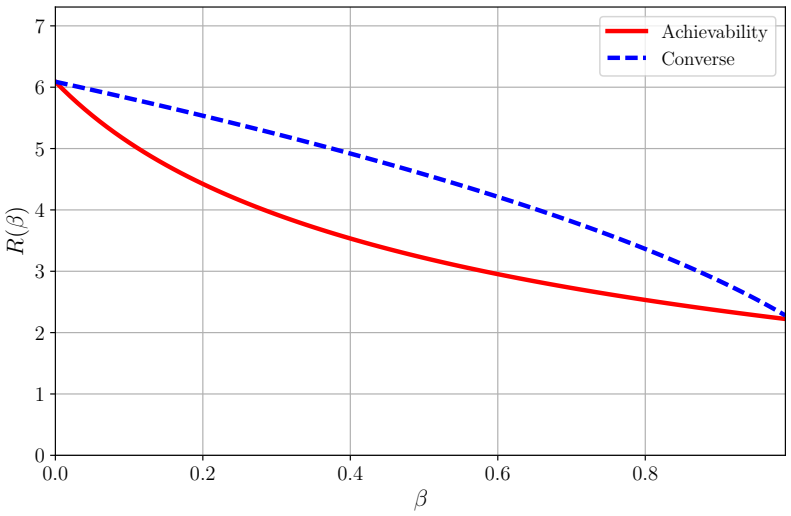

Figure 5.5.1: Illustration of Theorem 24 and Theorem 25 for a BSCs

In Fig. 5.5.1, we illustrate the result of Theorem 24 and 25 for the following example. When $s=0$, Bob and Willie's channel are independent BSC(0.1) and BSC(0.4), respectively. When $s=1$, Bob's channel is a binary asymmetric channel with flipping probability 0.1 and 0.2 for $x=0$ and $x=1$, respectively, while Willie's channel is BSC(0.3). The bounds are reasonably tight but, as expected, do not match.

5.5.1 Proof of Theorem 24

We break down the proof of Theorem 24 into six steps.

1. We first establish a technical lemma pertaining to a concentration inequality for the reciprocal of the sum of iid random variables.

2. We define an auxiliary problem and derive one-shot reliability and secrecy results.

3. We specialize the auxiliary problem to an arbitrarily varying DMC and assume that an oracle provides the weight of the warden's state sequence. We use the result of Step 2 to develop a universal secrecy and reliability scheme.

4. We reduce the amount of common randomness required for the coding scheme developed in Step 3.

5. We remove the oracle from the coding scheme by introducing estimators for the weight of the warden's state sequence.

6. Finally, we combine all steps to prove the result.

A Concentration Inequality

Suppose $\{X_i\}_{i=1}^n$ are iid according to Bernoulli(p). Since $\mathbb{E}(\sum_{i=1}^n X_i) = np$, one could expect $\mathbb{E}\left(\left|\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p}\right|\right)$ to be small. The following lemma formalizes this intuition and is proved in Appendix 5.B.

Lemma 42. *Suppose $X_1, \cdots, X_n$ are iid according to Bernoulli(p). Then, for $\frac{2}{np} < \epsilon < 1$,*

$$\mathbb{P}\left(\left|\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p}\right| \geqslant \frac{\epsilon}{(n+1)p}\right) \leqslant 2\exp\left(-\frac{np\epsilon^2}{32}\right), \tag{5.15}$$

$$\mathbb{E}\left(\left|\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p}\right|\right) \leqslant \frac{\epsilon}{(n+1)p} + \left(1 + \frac{1}{(n+1)p}\right) e^{-\frac{np\epsilon^2}{32}}. \tag{5.16}$$

One-shot Results for an Auxiliary Problem

We introduce an auxiliary problem with the help of which we solve the main problem later on. The main rationale for introducing the auxiliary problem is to use a variation of the likelihood encoder [77], which allows us to exploit channel coding tools. This helps us avoid a finite length penalty that would appear if using source coding tools and would dominate the covert throughput. Alice, Bob and Willie have access to $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$, respectively, with joint distribution P_{XYZ}. In addition, Alice and Bob share secret common randomness in the form of a random codebook $\widetilde{\mathbf{Y}} = \{\widetilde{Y}_{w_1 w_2}\}_{w_1 \in [\![1,M_1]\!], w_2 \in [\![1,M_2]\!]} \in \mathcal{Y}^{M_1 M_2}$ distributed according to $Q_Y^{\otimes M_1 M_2}$ for some chosen Q_Y. Bob generates two messages $W_1 \in [\![1, M_1]\!]$ and $W_2 \in [\![1, M_2]\!]$ from Y and $\widetilde{\mathbf{Y}}$ according to the conditional PMF

$$P_{W_1 W_2|Y\widetilde{\mathbf{Y}}}(w_1, w_2|y, \widetilde{\mathbf{y}}) = \begin{cases} \frac{\mathbb{1}\{y=\widetilde{y}_{w_1 w_2}\}}{\sum_{w_1' w_2'} \mathbb{1}\{y=\widetilde{y}_{w_1' w_2'}\}} & \sum_{w_1' w_2'} \mathbb{1}\{y = \widetilde{y}_{w_1' w_2'}\} \neq 0 \\ \frac{1}{M_1 M_2} & \sum_{w_1' w_2'} \mathbb{1}\{y = \widetilde{y}_{w_1' w_2'}\} = 0 \end{cases}. \tag{5.17}$$

This operation induces the joint distribution $P_{XYZ\widetilde{\mathbf{Y}}W_1W_2} \triangleq P_{XYZ} Q_Y^{\otimes M_1 M_2} P_{W_1W_2|Y\widetilde{\mathbf{Y}}}$. Note that Q_Y may be different from P_Y, which will be useful when analyzing universality and constitutes the main deviation from the standard likelihood encoder. The next two lemmas provide bounds showing that W_1 and W_2 can be interpreted as a secret key and a public message, respectively.

Lemma 43. *Let $\nu : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a fixed function and define a* universal *decoder for estimating W_1 from $X, \widetilde{\mathbf{Y}}, W_2$ as*

$$\hat{w}_1 = \phi(x, \widetilde{\mathbf{y}}, w_2) \triangleq \arg\max_{w_1} \nu\left(x, \widetilde{y}_{w_1 w_2}\right) \tag{5.18}$$

For all $\gamma > 0$, $\mu_Q \triangleq \min_y Q_Y(y)$, *and* $\frac{2}{(M_1M_2-1)\mu_Q} < \delta < 1$, *we have*

$$\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big) \leqslant \sum_{x,y} P_{XY}(x,y)\min(1, M_1 q(x,y)) + (2 + M_1M_2)e^{-\frac{(M_1M_2-1)\mu_Q\delta^2}{32}} + \delta, \quad (5.19)$$

where $q(x,y) \triangleq \sum_{y'} Q_Y(y')\mathbb{1}\{\nu(x,y') \geqslant \nu(x,y)\}$.

Proof. See Appendix 5.C. □

Lemma 44. *For all* $\gamma > 0$ *and* $\frac{2}{(M_1M_2-1)\mu_Q} < \delta < 1$, *we have*

$$\frac{1}{2}\big\|P_{W_1Z\widetilde{\mathbf{Y}}} - P_{W_1}^{\text{unif}} \otimes P_{Z\widetilde{\mathbf{Y}}}\big\|_1 \leqslant \sum_{y,z} P_{YZ}(y,z)\mathbb{1}\{P_{YZ}(y,z) \geqslant \gamma P_Z(z)Q_Y(y)\} + \frac{1}{2}\sqrt{\frac{\gamma}{M_2}} + \frac{1}{2}\delta + \frac{1}{2}(2 + M_1M_2)e^{-\frac{(M_1M_2-1)\mu_Q\delta^2}{32}}. \quad (5.20)$$

Proof. See Appendix 5.D. □

Universal Asymptotic Results for the Auxiliary Problem

We now extend the auxiliary problem results of Section 5.5.1 by using the channel n times allowing the warden to vary the channel at every channel use. More precisely, we consider an arbitrarily varying DMC $(\mathcal{X} \times \mathcal{S}, W_{YZ|XS}, \mathcal{Y}, \mathcal{Z})$ with $\mathcal{X} = \mathcal{S} = \{0,1\}$ and $P_0^0 = P_0^1 = P_0$. For simplicity, we suppose for now that the weight $\text{wt}(\mathbf{s})$ of Willie's state sequence is provided to Alice and Bob by some *oracle* at the end of the transmission. Alice samples the input sequence $\mathbf{X}$ according to $Q_X^{\otimes n}$ where $Q_X = \text{Bernoulli}(\alpha_n)$ and $\alpha_n \in \omega\left(\frac{\log n}{n}\right) \cap o\left(\frac{1}{\sqrt{n}}\right)$, and transmits it over the channel so that Bob and Willie observe $\mathbf{Y}$ and $\mathbf{Z}$, respectively. Alice and Bob are assumed to share a random codebook $\{\widetilde{\mathbf{Y}}_{w_1,w_2,w_3}\}_{w_1\in[\![1,M_1]\!],w_2\in[\![1,M_2]\!],w_3\in[\![1,M_3]\!]}$ distributed iid according to $P_0^{\otimes n}$. Bob randomly

creates an encoder F to generate W_1 and W_2 from $\mathbf{y}$ using $\{\mathbf{Y}_{w_1 w_2 w_3}\tilde{}\}$ as

$$\mathbb{P}(F(\mathbf{y}) = (w_1, w_2)) = \frac{\sum_{w_3} \mathbb{1}\left\{\mathbf{y} = \widetilde{\mathbf{Y}}_{w_1 w_2 w_3}\right\}}{\sum_{w_1' w_2' w_3'} \mathbb{1}\left\{\mathbf{y} = \widetilde{\mathbf{Y}}_{w_1' w_2' w_3'}\right\}} \tag{5.21}$$

if $\sum_{w_1' w_2' w_3'} \mathbb{1}\left\{\mathbf{y} = \widetilde{\mathbf{Y}}_{w_1' w_2' w_3'}\right\} \neq 0$ and $\frac{1}{M_1 M_2}$ else. Message W_2 is sent over the public channel and Alice subsequently uses $\mathbf{X}$ and W_2 to decode W_1 as $\widehat{W}_1$ with the random decoder $\Phi : \mathcal{X}^n \times [\![1, M_2]\!] \to [\![1, M_1]\!]$ defined as

$$\Phi(\mathbf{x}, w_2) \triangleq \arg\max_{w_1} \left(\max_{w_3} I(\mathbf{x} \wedge \widetilde{\mathbf{Y}}_{w_1 w_2 w_3}) \right). \tag{5.22}$$

For a code (f, ϕ) and a sequence of states $\mathbf{s}$, we define

$$P_e(f, \phi|\mathbf{s}) \triangleq \mathbb{P}\Big(W_1 \neq \widehat{W}_1 | \mathbf{s}\Big), \tag{5.23}$$

$$S(f, \phi|\mathbf{s}) \triangleq \frac{1}{2} \Big\| \widehat{P}_{W_1 W_2 \mathbf{Z}|\mathbf{s}} - \widehat{P}_{W_1 W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}} \Big\|_1. \tag{5.24}$$

We shall now prove that the above random code performs well over a class of state sequences specified as follows. For each $\mathbf{s}$, set $\beta \triangleq \frac{\mathrm{wt}(\mathbf{s})}{n}$, $Q_S \triangleq \text{Bernoulli}(\beta)$, and define the PMF

$$Q^{\beta}_{SXYZ}(s, x, y, z) \triangleq Q_S(s) Q_X(x) W_{YZ|XS}(y, z|x, s). \tag{5.25}$$

Define $\mathcal{S}(M_1, M_2, M_3)$ as the set of state sequences $\mathbf{s}$ such that the corresponding Q^{β}_{SXYZ} satisfies

$$\log M_1 + \log M_2 + \log M_3 \geqslant \lceil (1 + \zeta) \log \frac{1}{\mu_0} n \rceil, \tag{5.26}$$

$$\log M_1 + \log M_3 \leqslant \lfloor (1 - \zeta) \mathbb{I}(X; Y) n \rfloor, \tag{5.27}$$

$$\log M_3 \geqslant \lceil (1 + \zeta) \alpha_n \left(\beta I^1 + (1 - \beta) I^0 \right) n \rceil. \tag{5.28}$$

Note that W_1 should be interpreted as a secret key and W_2 should be interpreted as a public message. W_3 should be interpreted as additional randomization, which plays no role beyond helping us control the specific distribution used by the likelihood encoder. The next lemma shows that the random code described above is universal over the set $\mathcal{S}(M_1, M_2, M_3)$. Specifically, we prove that given the choice of (M_1, M_2, M_3), there exists a protocol that performs well for all state sequences $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$. In Section 5.5.1, where we choose the values of M_1, M_2, and M_3 based on Bob's observations, we show that with high probability the true state sequence is in $\mathcal{S}(M_1, M_2, M_3)$.

Lemma 45. *For all $\beta, \zeta > 0$, there exists $\xi > 0$ such that for large enough n and for all $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$, we have*

$$\mathbb{E}_{F,\Phi}(P_e(F,\Phi|\mathbf{s})) \leqslant 2^{-\omega(\log n)} \text{ and} \tag{5.29}$$

$$\mathbb{E}_{F,\Phi}(S(F,\Phi|\mathbf{s})) \leqslant 2^{-\xi\alpha_n n}, \tag{5.30}$$

where the term $\omega(\log n)$ depends on ζ and the channel.

Proof. See Appendix 5.E. □

Common Randomness Reduction

In the next lemma, we use Ahlswede's elimination technique [78] to reduce the amount of common randomness in the coding scheme.

Lemma 46. *Let (F, Φ) be any random code and $\mathcal{S}_n$ be a subset of $\mathcal{S}^n$. Furthermore, for all $\mathbf{s} \in \mathcal{S}_n$, assume that $\mathbb{E}_{F,\Phi}(P_e(F,\Phi|\mathbf{s})) \leqslant \epsilon$ and $\mathbb{E}_{F,\Phi}(S(F,\Phi|\mathbf{s})) \leqslant \epsilon$. Then, there exist L realizations $(f_1, \phi_1), \cdots, (f_L, \phi_L)$of the random code that satisfy*

$$\frac{1}{L}\sum_{i=1}^{L} P_e(f_i, \phi_i|\mathbf{s}) \leqslant \epsilon' \text{ and } \frac{1}{L}\sum_{i=1}^{L} S(f_i, \phi_i|\mathbf{s}) \leqslant \epsilon' \tag{5.31}$$

for all $\mathbf{s} \in \mathcal{S}_n$ *provided that*

$$\epsilon' > 2\log(1+\epsilon) \text{ and } L > \frac{2}{\epsilon'}(1+n). \tag{5.32}$$

Proof. Let $(F_1, \Phi_1), \cdots, (F_L, \Phi_L)$ be L iid random codes distributed according to $P_{(F,\Phi)}$. For any $\mathbf{s} \in \mathcal{S}_n$, we have

$$\begin{aligned}
&\mathbb{P}_{F,\Phi}\left(\frac{1}{L}\sum_{i=1}^{L} P_e(F_i,\Phi_i|\mathbf{s}) \geqslant \epsilon' \text{ or } \frac{1}{L}\sum_{i=1}^{L} S(F_i,\Phi_i|\mathbf{s}) \geqslant \epsilon'\right) && (5.33)\\
&\qquad \leqslant 2^{-L\epsilon'}\mathbb{E}\left(2^{\sum_{i=1}^{L} P_e(F_i,\Phi_i|\mathbf{s})}\right) + 2^{-L\epsilon'}\mathbb{E}\left(2^{\sum_{i=1}^{L} S(F_i,\Phi_i|\mathbf{s})}\right) && (5.34)\\
&\qquad = 2^{-L\epsilon'}\left(\mathbb{E}\left(2^{P_e(F,\Phi|\mathbf{s})}\right)\right)^L + 2^{-L\epsilon'}\left(\mathbb{E}\left(2^{S(F,\Phi|\mathbf{s})}\right)\right)^L && (5.35)\\
&\qquad \overset{(a)}{\leqslant} 2^{-L\epsilon'}\left(1+\mathbb{E}(P_e(F,\Phi|\mathbf{s}))\right)^L + 2^{-L\epsilon'}\left(1+\mathbb{E}(S(F,\Phi|\mathbf{s}))\right)^L && (5.36)\\
&\qquad \leqslant 2^{-L\epsilon'}\left(1+\epsilon\right)^L + 2^{-L\epsilon'}\left(1+\epsilon\right)^L && (5.37)\\
&\qquad = 2^{-L(\epsilon'-\log(1+\epsilon))+1}, && (5.38)
\end{aligned}$$

where (a) follows from $2^x \leqslant 1+x$ for $x \in [0,1]$. Therefore, the union bound yields that

$$\begin{aligned}
\mathbb{P}_{F,\Phi}\Bigg(&\forall \mathbf{s} \in \mathcal{S}_n, \frac{1}{L}\sum_{i=1}^{L} P_e(F_i,\Phi_i|\mathbf{s}) < \epsilon'\\
&\qquad \text{and } \frac{1}{L}\sum_{i=1}^{L} S(F_i,\Phi_i|\mathbf{s}) < \epsilon'\Bigg) > 1 - 2^n 2^{-L(\epsilon'-\log(1+\epsilon))+1}, \quad (5.39)
\end{aligned}$$

which is positive given that (5.32) holds. □

Corollary 2. *Under the same assumptions as Lemma 45, for all* $\zeta > 0$*, all large enough* n*, and* $L > 2n^4(1+n)$*, there exist codes* $(f_1,\phi_1), \cdots, (f_L,\phi_L)$ *such that for any* $\mathbf{s} \in \mathcal{S}_\beta(M_1,M_2,M_3)$*,*

$$\frac{1}{L}\sum_{\ell=1}^{L} P_e(f_\ell,\phi_\ell|\mathbf{s}) \leqslant \frac{1}{n^4}, \text{ and } \frac{1}{L}\sum_{\ell=1}^{L} S(f_\ell,\phi_\ell|\mathbf{s}) \leqslant \frac{1}{n^4}. \tag{5.40}$$

Proof. We first consider the random code (F, Φ) introduced in Section 5.5.1 for which we have $\mathbb{E}_{F,\Phi}(P_e(F,\Phi|\mathbf{s})) \leqslant 2^{-\omega(\log n)}$ and $\mathbb{E}_{F,\Phi}(S(F,\Phi|\mathbf{s})) \leqslant 2^{-\xi\alpha_n n} = 2^{-\omega(\log n)}$ by Lemma 45 for all $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$. Applying Lemma 46 to (F, Φ) for $L > 2n^4(1+n)$, we obtain L codes $(f_1, \phi_1), \cdots, (f_L, \phi_L)$ such that (5.40) holds since the two constraints $n^{-4} \geqslant 2\log(1 + 2^{-\omega(\log n)})$ and $L > \frac{2}{n^{-4}}(1+n)$ in Lemma 46 hold for large n. □

Estimation of the State Sequence Weight

We now construct an estimator for $\mathrm{wt}(\mathbf{s})$ to replace the oracle. This requires running the protocol over $n' = n + g$ channel uses, where g is a positive integer to be specified later used to provision for channel estimation. Before transmission, Alice and Bob secretly and independently select every channel use for estimation with probability $\kappa_n \in [0,1]$, which requires $n\mathbb{H}_b(\kappa_n)$ bits of shared secret key. Let L denote the number of positions chosen for the estimation and let $\mathbf{J} = (J_1, \cdots, J_L)$ denote the corresponding indices in increasing order. If $n' - n \triangleq g < L$, Alice and Bob halt the protocol and do not generate a key. Otherwise, Alice transmits symbol "1" in the positions in $\mathbf{J}$ and operates as in Step 3 in the known n positions not in $\mathbf{J}$. Since $P_0^1 \neq P_1^1$, there exists $y_0 \in \mathcal{Y}$ such that $P_0^1(y_0) \neq P_1^1(y_0)$. For $\mu_0 \triangleq P_0^1(y_0)$, $\mu_1 \triangleq P_1^1(y_0)$, and $T_i \triangleq \frac{\mathbb{1}\{Y_{j_i}=y_0\}-\mu_0}{\mu_1-\mu_0}$, Bob estimates $\beta = \frac{\mathrm{wt}(\mathbf{s})}{n}$ as $\widehat{\beta} \triangleq \frac{1}{L}\sum_{i=1}^{L} T_i$ (for $L = 0$, we define $\widehat{\beta} = 1$). Note that for a fixed j_i, $\mathbb{E}(T_i) = s_{j_i}$, and by the results on sampling without replacement, one can expect that $\sum_{i=1}^{L} T_i/L \approx \sum_{i=1}^{L} s_{J_i}/L \approx \sum_{i=1}^{n} s_i/n = \beta$.

We now show that, with high probability, Alice and Bob do not halt the protocol and $\widehat{\beta}$ is close to β. With $g = (1+\mu)\kappa_n n'$, application of a Chernoff bound yields that $\mathbb{P}(L \geqslant g) \leqslant 2^{-\frac{\mu^2\kappa_n n'}{3}}$. In addition, for all $\lambda > 0$ and $\mu \in]0,1[$,

$$\mathbb{P}\Big(|\widehat{\beta} - \beta| > \lambda\Big) \tag{5.41}$$

$$= \sum_{\ell=0}^{n'} \mathbb{P}(L = \ell)\mathbb{P}\Big(\Big|\widehat{\beta} - \beta\Big| > \lambda \Big| L = \ell\Big) \tag{5.42}$$

$$\leqslant \mathbb{P}(L \leqslant (1-\mu)\kappa_n n') + \sum_{\ell=\lfloor(1-\mu)\kappa_n n'\rfloor+1}^{n'} \mathbb{P}(L=\ell)\mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell\right) \tag{5.43}$$

$$\overset{(a)}{\leqslant} \exp\left(-\frac{1}{2}\mu^2\kappa_n n'\right) + \sum_{\ell=\lfloor(1-\mu)\kappa_n n'\rfloor+1}^{n'} \mathbb{P}(L=\ell)\mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell\right), \tag{5.44}$$

where (a) follows from a Chernoff bound. Conditioned on $L=\ell$, $\mathbf{J}=(J_1,\cdots,J_\ell)$ is distributed uniformly on $\mathcal{J}^\ell = \{\mathbf{j}=(j_1,\cdots,j_\ell): j_1<\cdots<j_\ell\}$. Upon defining the event

$$\mathcal{E} \triangleq \left\{\mathbf{j}\in\mathcal{J}^\ell : \left|\frac{1}{\ell}\sum_{i=1}^{\ell} s_{j_i} - \beta\right| > \frac{\lambda}{2}\right\}, \tag{5.45}$$

we have

$$\mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell\right) = \mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell, \mathcal{E}\right)\mathbb{P}(\mathcal{E}|L=\ell) \tag{5.46}$$

$$+ \mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell, \mathcal{E}^c\right)\mathbb{P}(\mathcal{E}^c|L=\ell) \tag{5.47}$$

$$\leqslant \mathbb{P}(\mathcal{E}|L=\ell) + \mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| > \lambda \middle| L=\ell, \mathcal{E}^c\right). \tag{5.48}$$

We next express $\mathbb{P}(\mathcal{E}|L=\ell)$ as the CDF of the hypergeometric distribution. In particular, let H denote the number of successes in ℓ draws without replacement from a population of size n' with $\mathrm{wt}(\mathbf{s})=\ell\beta$ successes in the population. We then have

$$\mathbb{P}(\mathcal{E}|L=\ell) = \mathbb{P}\left(\left|\frac{1}{\ell}H - \beta\right| \geqslant \frac{\lambda}{2}\right) \overset{(a)}{\leqslant} \exp\left(-\frac{\lambda^2\ell}{2}\right), \tag{5.49}$$

where (a) follows from the standard tail bounds for hypergeometric distribution (e.g., see

[79]).

We next fix some $\mathbf{j} \in \mathcal{J}^\ell \setminus \mathcal{E}$. Since $\mathbb{E}(T_i|J_i = j_i) = s_{j_i}$, and $\frac{-\mu_0}{\mu_1-\mu_0} \leqslant T_i \leqslant \frac{1-\mu_0}{\mu_1-\mu_0}$, Hoeffding's inequality (Theorem 3 in Chapter 1) implies that

$$\mathbb{P}\Big(|\widehat{\beta} - \beta| > \lambda | L = \ell, \mathbf{J} = \mathbf{j}\Big) \tag{5.50}$$

$$= \mathbb{P}\left(\left|\frac{1}{\ell}\sum_{i=1}^{\ell} T_i - \beta\right| \geqslant \lambda \middle| L = \ell, \mathbf{J} = \mathbf{j}\right) \tag{5.51}$$

$$\leqslant 2\exp\left(-2\ell(\mu_1-\mu_0)^2\left(\lambda - \left|\beta - \frac{1}{\ell}\sum_{i=1}^{\ell} s_{j_i}\right|\right)^2\right) \tag{5.52}$$

$$\overset{(a)}{\leqslant} 2\exp\left(-\frac{\ell(\mu_1-\mu_0)^2\lambda^2}{2}\right), \tag{5.53}$$

where (a) follows from $\mathbf{j} \notin \mathcal{E}$. Therefore, we obtain

$$\mathbb{P}\Big(|\widehat{\beta} - \beta| > \lambda | L = \ell\Big) \leqslant 2\exp\left(-\frac{(\mu_1-\mu_0)^2\lambda^2\ell}{2}\right) + 2\exp\left(-\frac{\lambda^2\ell}{2}\right), \tag{5.54}$$

which is less than $2^{-\xi\ell}$ for some $\xi > 0$ small enough and independent of ℓ. Combining (5.54) with (5.44), we obtain $\mathbb{P}\Big(|\widehat{\beta} - \beta| > \lambda\Big) \leqslant 2^{-\xi\kappa_n n'}$ for some $\xi > 0$ small enough.

We next show that for $Q_{\widehat{S}} = \text{Bernoulli}(\widehat{\beta})$, $Q^{\widehat{\beta}}_{\widehat{S}\widehat{X}\widehat{Y}\widehat{Z}}$ defined as in (5.25) and (M_1, M_2, M_3) such that

$$\log M_1 + \log M_2 + \log M_3 \geqslant \lceil (1+\zeta)\log\frac{1}{\mu_0} n \rceil, \tag{5.55}$$

$$\log M_1 + \log M_3 \leqslant \lfloor (1-\zeta)\left(\mathbb{I}(\widehat{X};\widehat{Y}) - \zeta\alpha_n\right) n \rfloor, \tag{5.56}$$

$$\log M_3 \geqslant \left\lceil (1+\zeta)\alpha_n\left(\widehat{\beta}I^1 + (1-\widehat{\beta})I^0 + \zeta\right) n \right\rceil, \tag{5.57}$$

we have $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$ with high probability. To do so, we should verify that (5.26)-

(5.28) hold with high probability. By our definition of (M_1, M_2, M_3), (5.26) is always true. Additionally, for $Q_S = \text{Bernoulli}(\beta)$, and Q^{β}_{SXYZ} defined as in (5.25), the function

$$\psi(\beta) \triangleq \mathbb{D}(\beta W_{Y|X=1S=1} + (1-\beta)W_{Y|X=1S=0} \\ \| \beta W_{Y|X=0S=1} + (1-\beta)W_{Y|X=0S=0}) \quad (5.58)$$

is continuous in β. Therefore, there exists $\lambda > 0$ such that if $|\beta - \beta'| < \lambda$, $|\psi(\beta) - \psi(\beta')| \leqslant \zeta - o(1)$. Then,

$$\begin{aligned} \mathbb{P}(\lfloor (1-\zeta)\left(\mathbb{I}(\widehat{X};\widehat{Y}) - \zeta\alpha_n\right) n \rfloor \geqslant \lfloor (1-\zeta)\mathbb{I}(X;Y)n \rfloor) \\ &= \mathbb{P}\left(\mathbb{I}(\widehat{X};\widehat{Y}) - \zeta\alpha_n \geqslant \mathbb{I}(X;Y) + O\left(\frac{1}{n}\right)\right) \\ &= \mathbb{P}\left(\alpha_n\psi(\widehat{\beta}) - \zeta\alpha_n \geqslant \alpha_n\psi(\beta) + O\left(\frac{1}{n}\right)\right) \\ &\leqslant \mathbb{P}\Big(|\psi(\beta) - \psi(\widehat{\beta})| \geqslant \zeta - o(1)\Big) \\ &\leqslant \mathbb{P}\Big(|\beta - \widehat{\beta}| \geqslant \lambda\Big) \\ &\leqslant 2^{-\xi\kappa_n n'}. \end{aligned} \quad (5.59)$$

Similarly, we can argue that

$$\mathbb{P}\left(\left\lceil (1+\zeta)\alpha_n\left(\widehat{\beta}I^1 + (1-\widehat{\beta})I^0 + \zeta\right) n\right\rceil \right. \\ \left. \leqslant \left\lceil (1+\zeta)\alpha_n\left(\beta I^1 + (1-\beta)I^0\right) n\right\rceil\right) \leqslant 2^{-\xi\kappa_n n'}, \quad (5.60)$$

so that

$$\mathbb{P}\left(\log M_1 \leqslant (1-\zeta)\left(\mathbb{I}(X;Y) - \zeta\alpha_n\right) n - (1+\zeta) \right. \\ \left. \times\alpha_n\left(\beta I^1 + (1-\beta)I^0 + \zeta\right) n - \zeta\alpha_n n\right) \leqslant 2^{-\xi\kappa_n n'}. \quad (5.61)$$

Hence, $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$ with probability more than $1 - 2^{-\xi\kappa_n n' + 1}$.

Proof of Theorem 24

We put together the different pieces developed so far and describe our active covert key generation protocol. Let $\zeta > 0$, $n' = n + g$ be the block-length, $\kappa_n = o(\alpha_n/\log n) \cap \omega(\log n/n)$, $g \geqslant (1+\mu)\kappa_n n'$ for some $\mu \in]0,1[$, $\mathbf{J} = (J_1, \cdots, J_L)$ be the positions to be used for the estimation, and K be a shared secret key uniformly distributed over $[\![1, U]\!]$ for any $U > 2n^4(n+1)$. For $L > g$, the protocol halts. Otherwise, Alice samples $\widetilde{\mathbf{X}}$ according to $Q_X^{\otimes n}$ for transmission over the channel $W_{YZ|XS}$ at the n positions not included in $\mathbf{J}$, and transmits 1 in the positions in $\mathbf{J}$. Let $\mathbf{Y}$ and $\mathbf{Z}$ denote Bob's and Willie's received sequences, respectively, and $\widetilde{\mathbf{Y}}$ denote the sub-sequence of $\mathbf{Y}$ obtained by removing the components in $\mathbf{J}$. Bob first estimates the type $\widehat{\beta}$ of Willie's states sequence defined in Section 5.5.1 and sets (M_1, M_2, M_3) such that (5.55)-(5.57) hold. Subsequently, for $(f_1, \phi_1), \cdots, (f_U, \phi_U)$ defined in Corollary 2, Bob generates two messages $(W_1, W_2) = f_K(\widetilde{\mathbf{Y}})$ and broadcasts W_2 together with $\widehat{\beta}$ one-time-padded with a shared secret key. Finally, Alice decodes W_1 as $\widehat{W_1} \triangleq \phi_K(\widetilde{\mathbf{X}}, W_2)$. We provide the performance analysis of the protocol in four parts.

Reliability Analysis With probability at most $2^{-\xi\kappa_n n'} \leqslant 2^{-\omega(\log n)}$, the protocol is halted. If $\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)$, then by Corollary 2, the probability of error is less than $\frac{1}{n^4}$. Since $\mathbb{P}(\mathbf{s} \in \mathcal{S}(M_1, M_2, M_3)) \geqslant 1 - 2^{-\xi\kappa_n n'+1} \leqslant 2^{-\omega(\log n)}$, the probability of error for the protocol is less than $n^{-4} + 2^{-\omega(\log n)}$.

Secrecy and Covertness Analysis Let $\widehat{P}_{W_1W_2\mathbf{Z}}$ be the PMF induced by the protocol and $\chi_2(\beta) \triangleq \beta\chi_2(Q_1^1 \| Q_0^1) + (1-\beta)\chi_2(Q_1^0 \| Q_0^0)$. By definition, we have

$$S(\mathcal{C}|\mathbf{s}) + C(\mathcal{C}|\mathbf{s}) \tag{5.62}$$

$$= \mathbb{D}\Big(\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} \| P_{W_1W_2}^{\text{unif}} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big) + \mathbb{D}\Big(\widehat{P}_{\mathbf{Z}|\mathbf{s}} \| Q_0^{\otimes n}\Big) \tag{5.63}$$

$$= \mathbb{D}\Big(\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} \| P_{W_1W_2}^{\text{unif}} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big) + \frac{1}{2}(\alpha_n + \kappa_n)^2\chi_2(\beta)n + O((\alpha_n + \kappa_n)^3 n) \tag{5.64}$$

$$\stackrel{(a)}{=} \mathbb{D}\Big(\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} \| P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big) + \frac{1}{2}\alpha_n^2\chi_2(\beta)n + o(\alpha_n^2 n) \tag{5.65}$$

$$\stackrel{(b)}{\leqslant} \frac{1}{2}\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1 \log(M_1M_2) + \mathbb{H}_b\left(\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1\right) + \frac{1}{2}\alpha^2\chi_2(\beta)n + o(\alpha_n^2 n) \tag{5.66}$$

$$\stackrel{(c)}{\leqslant} \frac{1}{2}\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1 \left(O(n) + \log \frac{e}{\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1}\right) + \frac{1}{2}\alpha_n^2\chi_2(\beta)n + o(\alpha_n^2 n). \tag{5.67}$$

where (a) follows since $\kappa_n = o(\alpha_n)$, (b) follows from [76, Problem 17.1], and (c) follows from $\mathbb{H}_b(x) \leqslant x \log \frac{e}{x}$. To upper-bound $\frac{1}{2}\Big\|\widehat{P}_{W_1W_3\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_3} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1$, let $\mathcal{E}$ be the event $\{\mathbf{s} \notin \mathcal{S}(M_1, M_2, M_3)\}$. By convexity of variational distance, we have

$$\frac{1}{2}\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathbf{s}}\Big\|_1 \leqslant \frac{1}{2}\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathcal{E}\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathcal{E}\mathbf{s}}\Big\|_1 \mathbb{P}(\mathcal{E}) \tag{5.68}$$

$$+ \frac{1}{2}\Big\|\widehat{P}_{W_1W_2\mathbf{Z}|\mathcal{E}^c\mathbf{s}} - P^{\text{unif}}_{W_1W_2} \otimes \widehat{P}_{\mathbf{Z}|\mathcal{E}^c\mathbf{s}}\Big\|_1 \mathbb{P}(\mathcal{E}^c) \tag{5.69}$$

$$\leqslant n^{-4} + 2^{-\xi\kappa_n n' + 1} \tag{5.70}$$

$$\leqslant n^{-4} + 2^{-\omega(\log n)}. \tag{5.71}$$

Hence, for large enough n, we have $S(\mathcal{C}|\mathbf{s}) + C(\mathcal{C}|\mathbf{s}) \leqslant n^{-2} + \frac{1}{2}\alpha_n^2\chi_2(\beta)n + o(\alpha_n^2 n)$, which is vanishing.

Rate Analysis The covert rate of the protocol is

$$\frac{\log M_1}{\sqrt{nC(\mathcal{C}|\mathbf{s})}} \geqslant \frac{\log M_1}{\sqrt{n\left(n^{-2} + \frac{1}{2}\alpha_n^2\chi_2(\beta)n + o(\alpha_n^2 n)\right)}}. \tag{5.72}$$

Moreover, by (5.61), with probability at least $2^{-\omega(\log n)}$, we have

$$\frac{\log M_1}{\sqrt{n\left(n^{-2}+\frac{1}{2}\alpha_n^2\chi_2(\beta)n+o(\alpha_n^2 n)\right)}} \tag{5.73}$$

$$\geqslant \frac{(1-\zeta)\left(\mathbb{I}(X;Y)-\zeta\alpha_n\right)n-(1+\zeta)\alpha_n\left(\beta I^1+(1-\beta)I^0+\zeta\right)n-\zeta\alpha_n n}{\sqrt{n\left(n^{-2}+\frac{1}{2}\alpha_n^2\chi_2(\beta)n+o(\alpha_n^2 n)\right)}} \tag{5.74}$$

$$=\sqrt{2}\frac{\mathbb{D}((1-\beta)P_1^0+\beta P_1^1\|P_0)-((1-\beta)I^0+\beta I^1)}{\sqrt{\chi_2(\beta)}}-o(1)-O(\zeta). \tag{5.75}$$

Finally, the required amount of secret common randomness for sharing K, sharing $\mathbf{J}$, and one-time-padding $\widehat{\beta}$ is $\log U = O(\log n)$, $n'\mathbb{H}_b\left(\kappa_n\right) = O(n'\kappa_n \log\frac{1}{\kappa_n}) = o(n'\alpha_n)$, and $O(\log n)$, respectively. Since all three terms are $o(n'\alpha_n)$, and the amount of the generated key is $\Omega(n'\alpha_n)$, the amount of secret common randomness is negligible.

APPENDIX

5.A Notation for Method of Types

For $\mathbf{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ and $a \in \mathcal{X}$, let $N(\mathbf{x}|a) \triangleq |\{i : x_i = a\}|$. If P_X is a PMF over $\mathcal{X}$, let $\mathcal{T}_{P_X} \triangleq \{\mathbf{x} \in \mathcal{X}^n : \text{for all } a \in \mathcal{X} :\ N(\mathbf{x}|a) = P(a)n\}$. We denote by $\mathcal{P}_n(\mathcal{X})$ the set of all pPMF P_X for which $\mathcal{T}_{P_X} \neq \emptyset$ and by $\mathcal{P}_n(\mathcal{X}|\mathcal{Y})$ the set of all conditional PMFs $P_{X|Y}$ for which there exists a joint PMF P_{XY} such that $P_{X|Y} = \frac{P_{XY}}{P_Y}$ and $\mathcal{T}_{P_{XY}} \neq 0$. For $\mathbf{x} \in \mathcal{X}^n$ and a conditional PMF $P_{Y|X}$, we also define $\mathcal{T}_{P_{Y|X}}(\mathbf{x}) \triangleq \{\mathbf{y} \in \mathcal{Y}^n : \text{for all } a \in \mathcal{X}, b \in \mathcal{Y} : N(\mathbf{x}, \mathbf{y}|a, b) = P_{Y|X}(b|a)N(\mathbf{x}|a)\}$. For two sequences $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$ such that $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{P_{XY}}$, we define $I(\mathbf{x} \wedge \mathbf{y}) \triangleq I(P_{XY})$.

5.B Proof of Lemma 42

We first use the additivity of probability measures for disjoint events to split the probability into two parts, i.e.,

$$\mathbb{P}\left(\left|\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p}\right| \geqslant \frac{\epsilon}{(n+1)p}\right) \tag{5.76}$$

$$= \mathbb{P}\left(\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p} \geqslant \frac{\epsilon}{(n+1)p}\right) + \mathbb{P}\left(\frac{1}{1+\sum_{i=1}^n X_i} - \frac{1}{(n+1)p} \leqslant -\frac{\epsilon}{(n+1)p}\right) \tag{5.77}$$

$$= \mathbb{P}\left(\sum_{i=1}^n X_i \leqslant \frac{(n+1)p}{1+\epsilon} - 1\right) + \mathbb{P}\left(\sum_{i=1}^n X_i \geqslant \frac{(n+1)p}{1-\epsilon} - 1\right) \tag{5.78}$$

$$\leqslant \mathbb{P}\left(\sum_{i=1}^n X_i \leqslant \frac{(n+1)p}{1+\epsilon}\right) + \mathbb{P}\left(\sum_{i=1}^n X_i \geqslant \frac{np}{1-\epsilon} - 1\right) \tag{5.79}$$

$$= \mathbb{P}\left(\sum_{i=1}^n X_i \leqslant \left(1 - \frac{\epsilon - \frac{1}{n}}{1+\epsilon}\right)np\right) + \mathbb{P}\left(\sum_{i=1}^n X_i \geqslant \left(1 + \frac{\epsilon}{1-\epsilon} - \frac{1}{np}\right)np\right) \tag{5.80}$$

$$\overset{(a)}{\leqslant} \mathbb{P}\left(\sum_{i=1}^n X_i \leqslant \left(1 - \frac{\epsilon}{2(1+\epsilon)}\right)np\right) + \mathbb{P}\left(\sum_{i=1}^n X_i \geqslant \left(1 + \frac{\epsilon}{2(1-\epsilon)}\right)np\right) \tag{5.81}$$

$$\overset{(b)}{\leqslant} \mathbb{P}\left(\sum_{i=1}^{n} X_i \leqslant \left(1-\frac{\epsilon}{4}\right)np\right) + \mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant \left(1+\frac{\epsilon}{2}\right)np\right), \tag{5.82}$$

where (a) follows since $\epsilon > \frac{2}{np}$, and (b) follows since $\epsilon \in [0,1]$. To upper-bound the above terms, we use known Chernoff bounds [80, Exercise 2.10] stating that for $\mu \in]0,1[$, we have

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \leqslant (1-\mu)np\right) \leqslant \exp\left(-\frac{np\mu^2}{2}\right), \tag{5.83}$$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant (1+\mu)np\right) \leqslant \exp\left(-\frac{np\mu^2}{3}\right). \tag{5.84}$$

Therefore, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \leqslant \left(1-\frac{\epsilon}{4}\right)np\right) \leqslant \exp\left(-\frac{np\epsilon^2}{32}\right), \tag{5.85}$$

and

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant \left(1+\frac{\epsilon}{2}\right)np\right) \leqslant \exp\left(-\frac{np\epsilon^2}{12}\right). \tag{5.86}$$

Combining these two inequalities completes the proof of (5.15).

To prove (5.16), we first define the event $\mathcal{E} \triangleq \left\{\left|\frac{1}{1+\sum_{i=1}^{n} X_i} - \frac{1}{(n+1)p}\right| \geqslant \frac{\epsilon}{(n+1)p}\right\}$. By the law of total probability,

$$\mathbb{E}\left(\left|\frac{1}{1+\sum_{i=1}^{n} X_i} - \frac{1}{(n+1)p}\right|\right) \tag{5.87}$$

$$= \mathbb{E}\left(\left|\frac{1}{1+\sum_{i=1}^{n} X_i} - \frac{1}{(n+1)p}\right| |\mathcal{E}\right)\mathbb{P}(\mathcal{E}) + \mathbb{E}\left(\left|\frac{1}{1+\sum_{i=1}^{n} X_i} - \frac{1}{(n+1)p}\right| |\mathcal{E}^c\right)\mathbb{P}(\mathcal{E}^c) \tag{5.88}$$

$$\leqslant \left(1+\frac{1}{(n+1)p}\right)e^{-\frac{np\epsilon^2}{32}} + \frac{\epsilon}{(n+1)p}. \tag{5.89}$$

5.C Proof of Lemma 43

By definition of our universal decoder and the construction of messages W_1 and W_2, we have

$$\begin{aligned}
\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big) &= \sum_{x,y,\widetilde{\mathbf{y}},w_1,w_2} P_{XY}(x,y) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) P_{W_1W_2|Y\widetilde{\mathbf{Y}}}(w_1,w_2|y,\widetilde{\mathbf{y}}) \\
&\qquad \times \mathbb{1}\big\{\exists w_1'' \neq w_1 :\ \nu(x,\widetilde{y}_{w_1'' w_2}) \geqslant \nu(x,\widetilde{y}_{w_1 w_2})\big\} \quad (5.90) \\
&\overset{(a)}{=} M_1 M_2 \sum_{x,y,\widetilde{\mathbf{y}}} P_{XY}(x,y) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) P_{W_1W_2|Y\widetilde{\mathbf{Y}}}(1,1|y,\widetilde{\mathbf{y}}) \\
&\qquad \times \mathbb{1}\big\{\exists w_1'' \neq 1 :\ \nu(x,\widetilde{y}_{w_1'' 1}) \geqslant \nu(x,\widetilde{y}_{11})\big\} \quad (5.91) \\
&\overset{(b)}{\leqslant} M_1 M_2 \sum_{x,y,\widetilde{\mathbf{y}}:\sum_{w_1' w_2'} \mathbb{1}\{y=\widetilde{y}_{w_1' w_2'}\}\neq 0} P_{XY}(x,y) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \\
&\qquad \times \frac{\mathbb{1}\{y=\widetilde{y}_{11}\}}{\sum_{w_1' w_2'} \mathbb{1}\{y=\widetilde{y}_{w_1' w_2'}\}} \mathbb{1}\big\{\exists w_1'' \neq 1 :\ \nu(x,\widetilde{y}_{w_1'' 1}) \geqslant \nu(x,\widetilde{y}_{11})\big\} \\
&\qquad + \sum_y P_Y(y)(1-Q_Y(y))^{M_1 M_2} \quad (5.92) \\
&\overset{(c)}{=} M_1 M_2 \sum_{x,\widetilde{\mathbf{y}}} P_{XY}(x,\widetilde{y}_{11}) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \\
&\times \frac{\mathbb{1}\big\{\exists w_1'' \neq 1 :\ \nu(x,\widetilde{y}_{w_1'' 1}) \geqslant \nu(x,\widetilde{y}_{11})\big\}}{\sum_{w_1' w_2'} \mathbb{1}\{\widetilde{y}_{11}=\widetilde{y}_{w_1' w_2'}\}} + \sum_y P_Y(y)(1-Q_Y(y))^{M_1 M_2}, \\
&\qquad (5.93)
\end{aligned}$$

where (a) follows since $Q_Y^{\otimes M_1 M_2}$ is iid, (b) follows since the probability that for all w_1' and w_2', we have $Y \neq \widetilde{Y}_{w_1' w_2'}$ is $\sum_y P_Y(y)(1-Q_Y(y))^{M_1 M_2}$, and (c) follows since we can replace y by $\widetilde{y}_{11}$ because of the term $\mathbb{1}\{y=\widetilde{y}_{11}\}$. We upper-bound the first term in (5.93) in two steps. Using Lemma 42, we first bound the difference between (5.93) and the same expressions when replacing $\sum_{w_1' w_2'} \mathbb{1}\{\widetilde{y}_{11}=\widetilde{y}_{w_1' w_2'}\}$ by its expected value $M_1 M_2 Q_Y(y)$. This follows from

$$
\begin{aligned}
&\left| \sum_{x,\widetilde{\mathbf{y}}} P_{XY}(x,\widetilde{y}_{11}) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \mathbb{1}\left\{\exists w_1'' \neq 1 :\ \nu(x,\widetilde{y}_{w_1''1}) \geqslant \nu(x,\widetilde{y}_{11})\right\} \right. \\
&\qquad\qquad \left. \times \left(\frac{1}{M_1 M_2 Q_Y(\widetilde{y}_{11})} - \frac{1}{\sum_{w_1' w_2'} \mathbb{1}\left\{\widetilde{y}_{11} = \widetilde{y}_{w_1' w_2'}\right\}} \right) \right| \\
&\leqslant \sum_{x,\widetilde{\mathbf{y}}} P_{XY}(x,\widetilde{y}_{11}) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \left| \frac{1}{M_1 M_2 Q_Y(\widetilde{y}_{11})} - \frac{1}{\sum_{w_1' w_2'} \mathbb{1}\left\{\widetilde{y}_{11} = \widetilde{y}_{w_1' w_2'}\right\}} \right| \\
&= \sum_{\widetilde{\mathbf{y}}} P_Y(\widetilde{y}_{11}) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \left| \frac{1}{M_1 M_2 Q_Y(\widetilde{y}_{11})} - \frac{1}{\sum_{w_1' w_2'} \mathbb{1}\left\{\widetilde{y}_{11} = \widetilde{y}_{w_1' w_2'}\right\}} \right| \\
&\overset{(a)}{\leqslant} \sum_{\widetilde{y}_{11}} P_Y(\widetilde{y}_{11}) Q_Y(\widetilde{y}_{11}) \left(\frac{\delta}{M_1 M_2 Q_Y(\widetilde{y}_{11})} + \left(1 + \frac{1}{M_1 M_2 Q_Y(\widetilde{y}_{11})}\right) e^{-\frac{(M_1 M_2 - 1) Q_Y(\widetilde{y}_{11}) \delta^2}{32}} \right) \\
&= \sum_{\widetilde{y}_{11}} P_Y(\widetilde{y}_{11}) \left(\frac{\delta}{M_1 M_2} + \left(Q_Y(\widetilde{y}_{11}) + \frac{1}{M_1 M_2}\right) e^{-\frac{(M_1 M_2 - 1) Q_Y(\widetilde{y}_{11}) \delta^2}{32}} \right) \\
&\leqslant \frac{1}{M_1 M_2} \left(\delta + (1 + M_1 M_2) e^{-\frac{(M_1 M_2 - 1) \mu_Q \delta^2}{32}} \right),
\end{aligned}
\tag{5.94}
$$

where (a) follows by applying Lemma 42 when $\widetilde{y}_{11}$ is fixed and other components of $\widetilde{\mathbf{Y}}$ are iid according to Q_Y. We now upper-bound

$$
\sum_{x,\widetilde{\mathbf{y}}} P_{XY}(x,\widetilde{y}_{11}) Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \frac{\mathbb{1}\left\{\exists w_1'' \neq 1 :\ \nu(x,y_{w_1''1}) \geqslant \nu(x,y_{11})\right\}}{Q_Y(\widetilde{y}_{11})} \tag{5.95}
$$

$$
= \sum_{x,\widetilde{y}_{11}} P_{XY}(x,\widetilde{y}_{11}) \sum_{\widetilde{\mathbf{y}} \setminus \{\widetilde{y}_{11}\}} Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \frac{\mathbb{1}\left\{\exists w_1'' \neq 1 :\ \nu(x,y_{w_1''1}) \geqslant \nu(x,y_{11})\right\}}{Q_Y(\widetilde{y}_{11})} \tag{5.96}
$$

$$
\leqslant \sum_{x,\widetilde{y}_{11}} P_{XY}(x,\widetilde{y}_{11}) \min\left(1, \sum_{\widetilde{\mathbf{y}} \setminus \{\widetilde{y}_{11}\}} Q_Y^{\otimes M_1 M_2}(\widetilde{\mathbf{y}}) \frac{\sum_{w_1'' \neq 1} \mathbb{1}\left\{\ \nu(x,y_{w_1''1}) \geqslant \nu(x,y_{11})\right\}}{Q_Y(\widetilde{y}_{11})}\right) \tag{5.97}
$$

$$
\leqslant \sum_{x,y} P_{XY}(x,y) \min(1, M_1 q(x,y)). \tag{5.98}
$$

Finally, to simplify our upper-bound on the average probability of error, we bound the second term in (5.93) as

$$\sum_y P_Y(y)(1-Q_Y(y))^{M_1M_2} \leqslant (1-\mu_Q)^{M_1M_2} \tag{5.99}$$

$$= e^{\ln(1-\mu_Q)M_1M_2} \tag{5.100}$$

$$\leqslant e^{-\frac{\mu_Q}{1-\mu_Q}M_1M_2} \tag{5.101}$$

$$\leqslant e^{-\frac{1}{32}\mu_Q M_1M_2\delta^2}, \tag{5.102}$$

which can be combined with (5.94) to obtain the desired result.

5.D Proof of Lemma 44

To simplify our notation, we treat P_{W_1Z} as a random PMF depending on $\widetilde{\mathbf{Y}}$, i.e.,

$$P_{W_1Z}(w_1,z) \triangleq \sum_y P_{YZ}(y,z)\frac{\sum_{w_2}\mathbb{1}\left\{y=\widetilde{Y}_{w_1w_2}\right\}}{\sum_{w_1'w_2'}\mathbb{1}\left\{y=\widetilde{Y}_{w_1'w_2'}\right\}} \tag{5.103}$$

when $\sum_{w_1'w_2'}\mathbb{1}\left\{y=\widetilde{Y}_{w_1'w_2'}\right\} \neq 0$ and $P_{W_1Z}(w_1,z) \triangleq P_Z(z)\frac{1}{M_1}$ otherwise. We can then write

$$\frac{1}{2}\left\|P_{W_1Z\widetilde{\mathbf{Y}}} - P_{W_1}^{\text{unif}} \otimes P_{Z\widetilde{\mathbf{Y}}}\right\|_1 = \mathbb{E}_{\widetilde{\mathbf{Y}}}\left(\frac{1}{2}\left\|P_{W_1Z} - P_{W_1}^{\text{unif}} \otimes P_Z\right\|_1\right). \tag{5.104}$$

We first define $\overline{P}_{W_1Z}$ as

$$\overline{P}_{W_1|Y}(w_1|y) \triangleq \frac{1}{M_1M_2Q_Y(y)}\sum_{w_2}\mathbb{1}\left\{y=\widetilde{Y}_{w_1w_2}\right\}, \tag{5.105}$$

$$\overline{P}_{W_1Z}(w_1,z) \triangleq \sum_y P_{YZ}(y,z)\overline{P}_{W_1|Y}(w_1|y), \tag{5.106}$$

which is not necessarily a PMF because the sum over all (w_1, z) may be less than one. Note that

$$\mathbb{E}_{\widetilde{\mathbf{Y}}}\big(\|\overline{P}_{W_1Z} - P_{W_1Z}\|_1\big) = \sum_{w_1,z} \mathbb{E}_{\widetilde{\mathbf{Y}}}\left(\left|\sum_y P_{YZ}(y,z)\left(\overline{P}_{W_1|Y}(w_1|y) - P_{W_1|Y}(w_1|y)\right)\right|\right) \tag{5.107}$$

$$\leqslant \sum_{w_1,z,y} P_{YZ}(y,z)\mathbb{E}_{\widetilde{\mathbf{Y}}}\big(\big|\overline{P}_{W_1|Y}(w_1|y) - P_{W_1|Y}(w_1|y)\big|\big) \tag{5.108}$$

$$= \sum_{w_1,y} P_Y(y)\mathbb{E}_{\widetilde{\mathbf{Y}}}\big(\big|\overline{P}_{W_1|Y}(w_1|y) - P_{W_1|Y}(w_1|y)\big|\big) \tag{5.109}$$

$$\overset{(a)}{=} M_1 \sum_y P_Y(y)\mathbb{E}_{\widetilde{\mathbf{Y}}}\big(\big|\overline{P}_{W_1|Y}(1|y) - P_{W_1|Y}(1|y)\big|\big), \tag{5.110}$$

where (a) follows since $Q_Y^{\otimes M_1 M_2}$ is iid. We also have

$$\mathbb{E}_{\widetilde{\mathbf{Y}}}\big(\big|\overline{P}_{W_1|Y}(1|y) - P_{W_1|Y}(1|y)\big|\big) \tag{5.111}$$

$$\leqslant \sum_{w_2} \mathbb{E}_{\widetilde{\mathbf{Y}}}\left(\mathbb{1}\left\{y = \widetilde{Y}_{1w_2}\right\}\left|\frac{1}{M_1M_2Q_Y(y)} - \frac{1}{\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}}\right|\left|\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}\right.\right.$$

$$\left.\left. \neq 0\right)\mathbb{P}_{\widetilde{\mathbf{Y}}}\left(\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\} \neq 0\right) + \frac{1}{M_1}\mathbb{P}_{\widetilde{\mathbf{Y}}}\left(\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\} = 0\right)\right. \tag{5.112}$$

$$= M_2\mathbb{E}_{\widetilde{\mathbf{Y}}}\left(\mathbb{1}\left\{y = \widetilde{Y}_{11}\right\}\left|\frac{1}{M_1M_2Q_Y(y)} - \frac{1}{\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}}\right|\left|\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}\right.\right.$$

$$\left.\left. \neq 0\right)\mathbb{P}_{\widetilde{\mathbf{Y}}}\left(\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\} \neq 0\right) + \frac{1}{M_1}\mathbb{P}_{\widetilde{\mathbf{Y}}}\left(\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\} = 0\right)\right. \tag{5.113}$$

$$= M_2Q_Y(y)\mathbb{E}_{\widetilde{\mathbf{Y}}\setminus\{\widetilde{Y}_{11}\}}\left(\left|\frac{1}{M_1M_2Q_Y(y)} - \frac{1}{1 + \sum_{w_1'w_2' \neq (1,1)}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}}\right|\right)$$

$$+ \frac{1}{M_1}\mathbb{P}_{\widetilde{\mathbf{Y}}}\left(\sum_{w_1'w_2'}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\} = 0\right) \tag{5.114}$$

$$= M_2Q_Y(y)\mathbb{E}_{\widetilde{\mathbf{Y}}\setminus\{\widetilde{Y}_{11}\}}\left(\left|\frac{1}{M_1M_2Q_Y(y)} - \frac{1}{1 + \sum_{w_1'w_2' \neq (1,1)}\mathbb{1}\left\{y = \widetilde{Y}_{w_1'w_2'}\right\}}\right|\right)$$

$$+\frac{1}{M_1}(1-Q_Y(y))^{M_1M_2} \tag{5.115}$$

$$\overset{(a)}{\leqslant} M_2Q_Y(y)\mathbb{E}_{\widetilde{\mathbf{Y}}\setminus\{\widetilde{Y}_{11}\}}\left(\left|\frac{1}{M_1M_2Q_Y(y)}-\frac{1}{1+\sum_{w_1'w_2'\neq(1,1)}\mathbb{1}\left\{y=\widetilde{Y}_{w_1'w_2'}\right\}}\right|\right)$$
$$+\frac{1}{M_1}e^{-\frac{1}{32}\mu_QM_1M_2\delta^2}, \tag{5.116}$$

where the derivation of (a) is similar to that of (5.102). We can now use Lemma 42, for a particular y and $\frac{2}{M_1M_2Q_Y(y)}<\delta<1$ to obtain

$$\mathbb{E}\left(\left|\frac{1}{M_1M_2Q_Y(y)}-\frac{1}{1+\sum_{w_1'w_2'\neq(1,1)}\mathbb{1}\left\{y=\widetilde{Y}_{w_1'w_2'}\right\}}\right|\right) \tag{5.117}$$
$$\leqslant \frac{\delta}{M_1M_2Q_Y(y)}+\left(1+\frac{1}{M_1M_2Q_Y(y)}\right)e^{-\frac{(M_1M_2-1)Q_Y(y)\delta^2}{32}} \tag{5.118}$$
$$\leqslant \frac{\delta}{M_1M_2Q_Y(y)}+\left(1+\frac{1}{M_1M_2Q_Y(y)}\right)e^{-\frac{(M_1M_2-1)\mu_Q\delta^2}{32}}. \tag{5.119}$$

Therefore, we obtain

$$\mathbb{E}_{\widetilde{\mathbf{Y}}}\left(\|\overline{P}_{W_1Z}-P_{W_1Z}\|_1\right)\leqslant\delta+(1+M_1M_2)e^{-\frac{(M_1M_2-1)\mu_Q\delta^2}{32}}$$
$$+\sum_y P_Y(y)(1-Q_Y(y))^{M_1M_2}. \tag{5.120}$$

We next decompose $\overline{P}_{W_1Z}$ into two components and define

$$\overline{P}^1_{W_1Z}(w_1,z)\triangleq\sum_y P_{YZ}(y,z)\overline{P}_{W_1|Y}(w_1|y)\mathbb{1}\{P_{YZ}(y,z)\geqslant\gamma P_Z(z)Q_Y(y)\}, \tag{5.121}$$

$$\overline{P}^2_{W_1Z}(w_1,z)\triangleq\sum_y P_{YZ}(y,z)\overline{P}_{W_1|Y}(w_1|y)\mathbb{1}\{P_{YZ}(y,z)<\gamma P_Z(z)Q_Y(y)\}, \tag{5.122}$$

for which we upper-bound $\mathbb{E}\Big(\|\overline{P}^1_{W_1Z} - \mathbb{E}\Big(\overline{P}^1_{W_1Z}\Big)\|_1\Big)$ and $\mathbb{E}\Big(\|\overline{P}^2_{W_1Z} - \mathbb{E}\Big(\overline{P}^2_{W_1Z}\Big)\|_1\Big)$ as

$$\begin{aligned}
\mathbb{E}\Big(\|\overline{P}^1_{W_1Z} - \mathbb{E}\Big(\overline{P}^1_{W_1Z}\Big)\|_1\Big) &\leqslant \mathbb{E}\Big(\|\overline{P}^1_{W_1Z}\|_1\Big) + \|\mathbb{E}\Big(\overline{P}^1_{W_1Z}\Big)\|_1 && (5.123)\\
&= 2\sum_{w_1,z} \mathbb{E}\Big(\overline{P}^1(w_1,z)\Big) && (5.124)\\
&= 2\sum_{w_1,z} \mathbb{E}\left(\sum_y P_{YZ}(y,z) \sum_{w_2} \frac{\mathbb{1}\Big\{y = \widetilde{Y}_{w_1w_2}\Big\}}{M_1M_2Q_Y(y)}\right) && (5.125)\\
&= 2\sum_{y,z} P_{YZ}(y,z)\mathbb{1}\{P_{YZ}(y,z) \geqslant \gamma P_Z(z)Q_Y(y)\}\,, && (5.126)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\Big(\|\overline{P}^2_{W_1Z} - \mathbb{E}\Big(\overline{P}^2_{W_1Z}\Big)\|_1\Big) && (5.127)\\
&\quad\leqslant \sum_{w_1,z} \mathbb{E}\Big(\Big|\overline{P}^2_{W_1Z}(w_1,z) - \mathbb{E}\Big(\overline{P}^2(w_1,z)\Big)\Big|\Big) && (5.128)\\
&\quad\leqslant \sum_{w_1,z} \sqrt{\mathrm{Var}\Big(\overline{P}^2(w_1,z)\Big)} && (5.129)\\
&\quad\leqslant \sum_{w_1,z} \sqrt{\sum_y \frac{P_{YZ}(y,z)^2 M_2 Q_Y(y)}{M_1^2M_2^2Q_Y(y)^2}\mathbb{1}\{P_{YZ}(y,z) < \gamma P_Z(z)Q_Y(y)\}} && (5.130)\\
&\quad\leqslant \sum_{w_1,z} \sqrt{\sum_y \frac{P_{YZ}(y,z)\gamma P_Z(z)Q_Y(y) M_2 Q_Y(y)}{M_1^2M_2^2Q_Y(y)^2}} && (5.131)\\
&\quad= \sqrt{\frac{\gamma}{M_2}}. && (5.132)
\end{aligned}$$

Therefore, by the triangle inequality, we obtain

$$\begin{aligned}
&\mathbb{E}\big(\|\overline{P}_{W_1Z} - \mathbb{E}\big(\overline{P}_{W_1Z}\big)\|_1\big)\\
&\qquad\leqslant 2\sum_{y,z} P_{YZ}(y,z)\mathbb{1}\{P_{YZ}(y,z) \geqslant \gamma P_Z(z)Q_Y(y)\} + \sqrt{\frac{\gamma}{M_2}}. \quad (5.133)
\end{aligned}$$

Combining (5.120) and (5.133) and noting that $\mathbb{E}\big(\overline{P}_{W_1Z}\big) = P^{\mathrm{unif}}_{W_1} \otimes P_Z$ completes the proof.

5.E Proof of Lemma 45

Applying Lemma 43 to (W_1, W_3) and W_2 for a fixed $\mathbf{s}$, we obtain

$$\mathbb{E}_{F,\Phi}(P_e(F,\Phi|\mathbf{s})) \tag{5.134}$$

$$\leqslant \sum_{\mathbf{x},\mathbf{y}} Q_X^{\otimes n}(\mathbf{x}) W_{Y|XS}^{\otimes n}(\mathbf{y}|\mathbf{x}\mathbf{s}) \min(1, M_1 M_3 q(\mathbf{x},\mathbf{y})) + \delta + (2 + M_1 M_2 M_3) e^{-\frac{(M_1 M_2 M_3 - 1)\mu_0^n \delta^2}{32}} \tag{5.135}$$

$$= \sum_{\mathbf{x},\mathbf{y}} Q_{XY|S}^{\otimes n}(\mathbf{x},\mathbf{y}|\mathbf{s}) \min(1, M_1 M_3 q(\mathbf{x},\mathbf{y})) + \delta + (2 + M_1 M_2 M_3) e^{-\frac{(M_1 M_2 M_3 - 1)\mu_0^n \delta^2}{32}}, \tag{5.136}$$

where

$$q(\mathbf{x},\mathbf{y}) \triangleq \sum_{\widetilde{\mathbf{y}}} P_0^{\otimes n}(\widetilde{\mathbf{y}}) \mathbb{1}\{I(\mathbf{x} \wedge \mathbf{y}) \leqslant I(\mathbf{x} \wedge \widetilde{\mathbf{y}})\}. \tag{5.137}$$

To upper-bound $q(\mathbf{x},\mathbf{y})$, let V_X and $V_{Y|X}$ be the type of $\mathbf{x}$ and the conditional type of $\mathbf{y}$ given $\mathbf{x}$, respectively. Then, we have

$$\sum_{\widetilde{\mathbf{y}}} P_0^{\otimes n}(\widetilde{\mathbf{y}}) \mathbb{1}\{I(\mathbf{x} \wedge \mathbf{y}) \leqslant I(\mathbf{x} \wedge \widetilde{\mathbf{y}})\}] \tag{5.138}$$

$$= \sum_{\widetilde{V}_{Y|X}} P_0^{\otimes n}(\mathcal{T}_{\widetilde{V}_{Y|X}}(\mathbf{x})) \mathbb{1}\Big\{I(V_X, \widetilde{V}_{Y|X}) \geqslant I(V_X, V_{Y|X})\Big\} \tag{5.139}$$

$$\overset{(a)}{\leqslant} \sum_{\widetilde{V}_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|\mathcal{X})} 2^{-n\mathbb{D}\left(\widetilde{V}_{Y|X} \| P_0 | V_X\right)} \mathbb{1}\Big\{I(V_X, \widetilde{V}_{Y|X}) \geqslant I(V_X, V_{Y|X})\Big\} \tag{5.140}$$

$$\leqslant (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n \min_{\widetilde{V}_{Y|X}: I(V_X, \widetilde{V}_{Y|X}) \geqslant I(V_X, V_{Y|X})} \mathbb{D}\left(\widetilde{V}_{Y|X} \| P_0 | V_X\right)} \tag{5.141}$$

$$= (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n \min_{\widetilde{V}_{Y|X}: I(V_X, \widetilde{V}_{Y|X}) \geqslant I(V_X, V_{Y|X})} I(V_X, \widetilde{V}_{Y|X}) + \mathbb{D}\left(\widetilde{V}_{Y|X} \circ V_X \| P_0\right)} \tag{5.142}$$

$$\leqslant (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-n \min_{\widetilde{V}_{Y|X}: I(V_X, \widetilde{V}_{Y|X}) \geqslant I(V_X, V_{Y|X})} I(V_X, \widetilde{V}_{Y|X})} \tag{5.143}$$

$$= (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nI(V_X, V_{Y|X})} \tag{5.144}$$

$$= (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nI(V_{X|S} \circ Q_S, V_{Y|XS} \circ Q_S)}, \tag{5.145}$$

where (a) follows from [76, Lemma 2.6]. Substituting the above upper-bound in the first term of the right hand side of (5.136) and using $Q_{XY|S}^{\otimes n}(\mathcal{T}_{V_{XY|S}}(\mathbf{s})|\mathbf{s}) \leqslant 2^{-n\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right)}$ [76, Equation (2.8)], we have

$$\sum_{\mathbf{x},\mathbf{y}} Q_X^{\otimes n}(\mathbf{x}) W_{Y|XS}^{\otimes n}(\mathbf{y}|\mathbf{x}\mathbf{s}) \min(1, M_1 M_3 q(\mathbf{x},\mathbf{y})) \tag{5.146}$$

$$\leqslant \sum_{V_{XY|S} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}|\mathcal{S})} 2^{-n\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right)} \min\left(1, M_1 M_3 (n+1)^{|\mathcal{X}||\mathcal{Y}|} 2^{-nI(V_{X|S}\circ Q_S, V_{Y|XS}\circ Q_S)}\right) \tag{5.147}$$

$$= \sum_{V_{XY|S} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}|\mathcal{S})} 2^{-n\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right)} 2^{-n[I(V_{X|S}\circ Q_S, V_{Y|XS}\circ Q_S) - \log M_1 M_3/n - O(\log n)/n]^+} \tag{5.148}$$

$$\leqslant (n+1)^{2|\mathcal{X}||\mathcal{Y}|} 2^{-n \min_{V_{XY|S}} \left(\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right) + [I(V_{X|S}\circ Q_S, V_{Y|XS}\circ Q_S) - \log M_1 M_3/n - O(\log n)/n]^+\right)} \tag{5.149}$$

$$\leqslant 2^{-n \min_{V_{XY|S}} \left(\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right) + [I(V_{X|S}\circ Q_S, V_{Y|XS}\circ Q_S) - \log M_1 M_3/n]^+ - O(\log n)/n\right)} \tag{5.150}$$

$$\overset{(a)}{\leqslant} 2^{-n \min_{V_{XY}} \left(\mathbb{D}(V_{XY}\|Q_{XY}) + [I(V_X, V_{Y|X}) - \log M_1 M_3/n]^+ - O(\log n)/n\right)} \tag{5.151}$$

$$\overset{(b)}{\leqslant} 2^{-n \min_{V_{XY}} \left(\mathbb{D}(V_{XY}\|Q_{XY}) + [I(V_X, V_{Y|X}) - (1-\zeta) I(Q_X, Q_{Y|X})]^+ - O(\log n)/n\right)}, \tag{5.152}$$

where (a) follows because $\mathbb{D}\left(V_{XY|S}\|Q_{XY|S}|Q_S\right) \geqslant \mathbb{D}(V_{XY}\|Q_{XY})$ by convexity of the KL-divergence, and (b) follows from (5.27). We next state a result that shows that, for all V_{XY}, $\mathbb{D}(V_{XY}\|Q_{XY})$ and $[I(V_X, V_{Y|X}) - (1-\zeta) I(Q_X, Q_{Y|X})]^+$ cannot be simultaneously small.

Lemma 47. *For a PMF V_{XY} with $\mathbb{D}(V_{XY}\|Q_{XY}) \leqslant \epsilon$, we have*

$$I(V_X, V_{Y|X}) \geqslant \left(\alpha - \sqrt{2\epsilon\alpha}\right) \left(\mathbb{D}\left(Q_{Y|X=1}\|Q_{Y|X=0}\right) - \sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}} \left(B + \frac{1}{2}\log \frac{1}{\sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}}}\right)\right) - \frac{\left(\alpha + \sqrt{2\epsilon}\right)^2 |\mathcal{Y}|}{\mu_0 - \sqrt{\frac{\epsilon}{1-\widetilde{\alpha}}}},$$

where $\alpha \triangleq Q_X(1)$, $\mu_0 \triangleq \min_y Q_{Y|X}(y|0)$, *and* B *is a constant that depends only on* $|\mathcal{Y}|$ *and* μ_0.

Proof. See Appendix 5.F. □

To lower-bound the exponent in (5.152),

$$n \min_{V_{XY}} (\mathbb{D}(V_{XY}\|Q_{XY}) + [I(V_X, V_{Y|X}) - (1-\zeta)I(Q_X, Q_{Y|X})]^+ - O\left(\log n\right)/n), \tag{5.153}$$

we consider two cases for V_{XY}. For $\{\epsilon_n\}_{n\geqslant 1} = \omega\left(\frac{\log n}{n}\right) \cap o(\alpha_n)$ and $\mathbb{D}(V_{XY}\|Q_{XY}) \geqslant \epsilon_n$, we have

$$\begin{aligned}\mathbb{D}(V_{XY}\|Q_{XY}) + [I(V_X, V_{Y|X}) - (1-\zeta)I(Q_X, Q_{Y|X})]^+ - O\left(\frac{\log n}{n}\right) \\ \geqslant \epsilon_n + O\left(\frac{\log n}{n}\right) \overset{(a)}{=} \omega\left(\frac{\log n}{n}\right),\end{aligned} \tag{5.154}$$

where (a) follows since $\epsilon_n = \omega(\log n/n)$. For the case when $\mathbb{D}(V_{XY}\|Q_{XY}) \leqslant \epsilon_n$, applying Lemma 47, we obtain

$$\mathbb{D}(V_{XY}\|Q_{XY}) + [I(V_X, V_{Y|X}) - (1-\zeta)I(Q_X, Q_{Y|X})]^+ - O(\log n/n) \tag{5.155}$$

$$\geqslant [I(V_X, V_{Y|X}) - (1-\zeta)I(Q_X, Q_{Y|X})]^+ - O(\log n/n) \tag{5.156}$$

$$\geqslant \Bigg[\left(\alpha_n - \sqrt{2\epsilon_n\alpha_n}\right)\left(\mathbb{D}\big(Q_{Y|X=1}\|Q_{Y|X=0}\big) - \sqrt{\frac{\epsilon_n}{\alpha_n - \sqrt{2\epsilon_n\alpha_n}}}\left(B + \frac{1}{2}\right.\right.$$

$$\times \log \frac{1}{\sqrt{\frac{\epsilon_n}{\alpha_n - \sqrt{2\epsilon_n \alpha_n}}}} \Bigg) \Bigg) - \frac{\left(\alpha_n + \sqrt{2\epsilon_n}\right)^2 |\mathcal{Y}|}{\mu_0 - \sqrt{\frac{\epsilon_n}{1-\widetilde{\alpha_n}}}} - (1-\zeta) I(Q_X, Q_{Y|X}) \Bigg]^+ - O\left(\log n/n\right) \tag{5.157}$$

$$\stackrel{(a)}{=} \left[\alpha_n (1 - o(1)) \left(\mathbb{D}\big(Q_{Y|X=1} \| Q_{Y|X=0}\big) - o(1)\right) - o(\alpha_n) - (1-\zeta) I(Q_X, Q_{Y|X})\right]^+ - O(\log n/n) \tag{5.158}$$

$$= \left[\alpha_n \mathbb{D}\big(Q_{Y|X=1} \| Q_{Y|X=0}\big) - (1-\zeta) I(Q_X, Q_{Y|X}) - o(\alpha_n)\right]^+ - O(\log n/n) \tag{5.159}$$

$$\stackrel{(b)}{\geqslant} \left[\alpha_n \mathbb{D}\big(Q_{Y|X=1} \| Q_{Y|X=0}\big) - (1-\zeta)\alpha_n \mathbb{D}\big(Q_{Y|X=1} \| Q_{Y|X=0}\big) - o(\alpha_n)\right]^+ - O(\log n/n) \tag{5.160}$$

$$= (1 - o(1))\zeta \alpha_n \mathbb{D}\big(Q_{Y|X=1} \| Q_{Y|X=0}\big) \tag{5.161}$$

$$= \omega\left(\frac{\log n}{n}\right), \tag{5.162}$$

where (a) follows since $\sqrt{2\epsilon_n \alpha_n} = o(\alpha_n)$, $\sqrt{\frac{\epsilon_n}{\alpha_n - \sqrt{2\epsilon_n \alpha_n}}} = o(1)$, and $\left(\alpha_n + \sqrt{2\epsilon_n}\right)^2 = o(\alpha_n)$, and (b) follows from [13, Lemma 1]. Therefore, we conclude that

$$n \min_{V_{XY}} (\mathbb{D}(V_{XY} \| Q_{XY}) + [I(V_X, V_{Y|X}) - (1-\zeta) I(Q_X, Q_{Y|X})]^+ - O(\log n)/n) = \omega(\log n), \tag{5.163}$$

and

$$\sum_{\mathbf{x},\mathbf{y}} Q_X^{\otimes n}(\mathbf{x}) W_{Y|XS}^{\otimes n}(\mathbf{y}|\mathbf{x}\mathbf{s}) \min(1, M_1 M_3 q(\mathbf{x}, \mathbf{y})) \leqslant 2^{-\omega(\log n)}. \tag{5.164}$$

Furthermore, the choice $\delta = 2^{-\frac{1}{3}\zeta n \log \frac{1}{\mu_0}}$ satisfies $\frac{2}{M_1 M_2 M_3 \mu_0^n} \leqslant \delta < 1$ for large n and we

obtain

$$\delta + (2 + M_1M_2M_3)e^{-\frac{(M_1M_2M_3-1)\mu_0\delta^2}{32}} \leqslant \frac{2^{-\frac{1}{3}\zeta \log \frac{1}{\mu_0} n}}{1 - 2^{-\frac{1}{3}\zeta \log \frac{1}{\mu_0} n}} + \left(2 + 2^{\lceil (1+\zeta) n \log \frac{1}{\mu_0} \rceil}\right) e^{-\frac{2^{\frac{1}{2}\zeta \log \frac{1}{\mu_0} n}}{32}}, \quad (5.165)$$

which is less than $e^{-\zeta' n}$ for large enough n and $\zeta' > 0$ independent of n.

To analyze the secrecy, we fix $\mathbf{s} \in \mathcal{S}^n$ with $\mathrm{wt}(\mathbf{s}) = \beta n$ and define the PMF $P_{\mathbf{YZ}}(\mathbf{y}, \mathbf{z}) \triangleq \sum_{\mathbf{x}} Q_X^{\otimes n}(\mathbf{x}) W_{YZ|XS}^{\otimes n}(\mathbf{yz}|\mathbf{xs})$, by Lemma 44, we have

$$\mathbb{E}_{F,\Phi}(S(F, \Phi|\mathbf{s})) \leqslant \mathbb{P}_{P_{\mathbf{YZ}}}\left(\sum_{i=1}^{n} \log \frac{P_{Y_i|Z_i}(Y_i|Z_i)}{P_0(Y_i)} \geqslant \log \gamma\right) + \frac{1}{2}\sqrt{\frac{\gamma}{M_3}} + \frac{1}{2}\delta + \frac{1}{2}(2 + M_1M_2M_3)e^{-\frac{(M_1M_2M_3-1)\mu_0^n\delta^2}{32}}.$$

Moreover, for $t \triangleq \gamma - \sum_{i=1}^{n} \mathbb{E}\left(\log \frac{P_{Y_i|Z_i}(Y_i|Z_i)}{P_0(Y_i)}\right) \geqslant 0$ and

$$C \triangleq \max_{y,z} \left(\log \frac{\sum_x Q_X(x) W_{YZ|XS}(y, z|x, 0))}{P_0(y)}, \log \frac{\sum_x Q_X(x) W_{YZ|XS}(y, z|x, 1))}{P_0(y)}\right) = O(1), \quad (5.166)$$

Bernstein's inequality yields that for

$$\mathbb{P}_{P_{\mathbf{YZ}}}\left(\sum_{i=1}^{n} \log \frac{P_{Y_i|Z_i}(Y_i|Z_i)}{P_0(Y_i)} \geqslant \log \gamma\right) \leqslant \exp\left(\frac{t^2}{\sum_{i=1}^{n} \mathrm{Var}\left(\log \frac{P_{Y_i|Z_i}(Y_i|Z_i)}{P_0(Y_i)}\right) + \frac{1}{3}Ct}\right). \quad (5.167)$$

By [69, Lemma 2], we also have $\mathbb{E}_{P_{Y_iZ_i}}\left(\log \frac{P_{Y_i|Z_i}(Y_i|Z_i)}{P_0(Y_i)}\right) = \mathbb{I}(Y_i; Z_i) + \mathbb{D}(P_{Y_i}\|P_0) = \alpha_n I^{s_i} + O(\alpha_n^2)$, where I^s is defined in the statement of Theorem 24. Thus, by choosing $\gamma = (1 + \zeta/2)\alpha_n(\beta I^1 + (1 - \beta)I^0)$, we obtain that $\mathbb{E}_{F,\Phi}(S(F, \Phi|\mathbf{s})) \leqslant 2^{-\xi n}$ for some

$\xi > 0$ small enough.

5.F Proof of Lemma 47

For a fixed V_{XY} and Q_{XY}, we first define

$$\alpha \triangleq Q_X(1), \quad \widetilde{\alpha} \triangleq V_X(1), \tag{5.168}$$

$$P_0 \triangleq Q_{Y|X=0}, \quad \widetilde{P}_0 \triangleq V_{Y|X=0}, \tag{5.169}$$

$$P_1 \triangleq Q_{Y|X=1}, \quad \widetilde{P}_1 \triangleq V_{Y|X=1}, \tag{5.170}$$

$$P_s \triangleq sQ_{Y|X=1} + (1-s)Q_{Y|X=0}, \tag{5.171}$$

$$\widetilde{P}_s \triangleq sV_{Y|X=1} + (1-s)V_{Y|X=0}. \tag{5.172}$$

By [13, Lemma 1], we have

$$I(V_{XY}) = \widetilde{\alpha}\mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) - \mathbb{D}\Big(\widetilde{P}_{\widetilde{\alpha}}\|\widetilde{P}_0\Big) \tag{5.173}$$

$$\geqslant \widetilde{\alpha}\mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) - \widetilde{\alpha}^2\chi_2(\widetilde{P}_1\|\widetilde{P}_0). \tag{5.174}$$

Moreover, by the chain rule for relative entropy, we can write $\mathbb{D}(V_{XY}\|Q_{XY})$ as

$$\mathbb{D}(V_{XY}\|Q_{XY}) = \mathbb{D}(V_X\|Q_X) + \mathbb{D}\big(V_{Y|X}\|Q_{Y|X}|V_X\big) \tag{5.175}$$

$$= \mathbb{D}(\widetilde{\alpha}\|\alpha) + \widetilde{\alpha}\mathbb{D}\Big(\widetilde{P}_1\|P_1\Big) + (1-\widetilde{\alpha})\mathbb{D}\Big(\widetilde{P}_0\|P_0\Big), \tag{5.176}$$

where $\mathbb{D}(p\|q) \triangleq p\log(p/q) + (1-p)\log((1-p)/(1-q))$. Since all terms in terms in (5.176) are positive, our assumption that $\mathbb{D}(V_{XY}\|Q_{XY}) \leqslant \epsilon$ implies that

$$\mathbb{D}(\widetilde{\alpha}\|\alpha) \leqslant \epsilon, \tag{5.177}$$

$$\mathbb{D}\Big(\widetilde{P}_1\|P_1\Big) \leqslant \frac{\epsilon}{\widetilde{\alpha}}, \tag{5.178}$$

$$\mathbb{D}\Big(\widetilde{P}_0\|P_0\Big) \leqslant \frac{\epsilon}{1-\widetilde{\alpha}}. \tag{5.179}$$

Using the inequalities $\mathbb{D}(p\|q) \geqslant (p-q)^2/(2q)$ for $p \leqslant q$ and $\mathbb{D}(p\|q) \geqslant (p-q)^2/(2p)$ for $q \leqslant p$, we obtain

$$\alpha - \sqrt{2\epsilon\alpha} \leqslant \widetilde{\alpha} \leqslant \alpha + \sqrt{2\epsilon}. \tag{5.180}$$

Furthermore, Pinsker's inequality yields that $\frac{1}{2}\left\|\widetilde{P}_1 - P_1\right\|_1 \leqslant \sqrt{\frac{\epsilon}{\widetilde{\alpha}}}$ and $\frac{1}{2}\left\|\widetilde{P}_0 - P_0\right\|_1 \leqslant \sqrt{\frac{\epsilon}{1-\widetilde{\alpha}}}$. Hence,

$$\begin{aligned}
\mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) &= \mathbb{D}(P_1\|P_0) + \mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) - \mathbb{D}(P_1\|P_0) && (5.181)\\
&\geqslant \mathbb{D}(P_1\|P_0) - \sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}} \\
&\quad \times \left(\frac{\log(|\mathcal{Y}|-1)}{2} + |\mathcal{Y}|\log\frac{1}{\mu_0} + \frac{|\mathcal{Y}|^2}{\mu_0 - \sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}}\right) - \mathbb{H}_b\left(\frac{\sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}}{2}\right) && (5.182)\\
&\geqslant \mathbb{D}(P_1\|P_0) - \sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}}\left(\frac{\log(|\mathcal{Y}|-1)}{2} + |\mathcal{Y}|\right. \\
&\qquad \left. \times \log\frac{1}{\mu_0} + \frac{|\mathcal{Y}|^2}{\mu_0 - \sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}} + \frac{1}{2}\log\frac{2e}{\sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}}\right) && (5.183)\\
&= \mathbb{D}(P_1\|P_0) - \sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}}\left(B + \frac{1}{2}\log\frac{1}{\sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}}\right), && (5.184)
\end{aligned}$$

where B depends only on μ_0 and $|\mathcal{Y}|$. Also because $\chi_2(\widetilde{P}_1\|\widetilde{P}_0) \leqslant |\mathcal{Y}|\,\frac{1}{\widetilde{\mu}_0}$ and $\widetilde{\mu}_0 \geqslant \mu_0 - \sqrt{\frac{\epsilon}{1-\widetilde{\alpha}}}$, we have

$$\begin{aligned}
I(V_{XY}) &= \widetilde{\alpha}\mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) - \mathbb{D}\Big(\widetilde{P}_{\widetilde{\alpha}}\|\widetilde{P}_0\Big) && (5.185)\\
&\geqslant \widetilde{\alpha}\mathbb{D}\Big(\widetilde{P}_1\|\widetilde{P}_0\Big) - \widetilde{\alpha}^2\chi_2(\widetilde{P}_1\|\widetilde{P}_0) && (5.186)\\
&\geqslant \Big(\alpha - \sqrt{2\epsilon\alpha}\Big)\left(\mathbb{D}(P_1\|P_0) - \sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}\left(B + \frac{1}{2}\log\frac{1}{\sqrt{\frac{\epsilon}{\alpha-\sqrt{2\epsilon\alpha}}}}\right)\right)
\end{aligned}$$

$$- \left(\alpha + \sqrt{2\epsilon}\right)^2 |\mathcal{Y}| \frac{1}{\mu_0 - \sqrt{\frac{\epsilon}{1-\widetilde{\alpha}}}} \tag{5.187}$$

$$= \alpha \mathbb{D}(P_1 \| P_0) - \sqrt{2\epsilon\alpha}\mathbb{D}(P_1 \| P_0) - \sqrt{\epsilon\left(\alpha - \sqrt{2\epsilon\alpha}\right)} \left(B + \frac{1}{2} \log \frac{1}{\sqrt{\frac{\epsilon}{\alpha - \sqrt{2\epsilon\alpha}}}} \right)$$
$$- \left(\alpha + \sqrt{2\epsilon}\right)^2 |\mathcal{Y}| \frac{1}{\mu_0 - \sqrt{\frac{\epsilon}{1-\widetilde{\alpha}}}}. \tag{5.188}$$

CHAPTER 6

COVERT SECRET KEY GENERATION: CLASSICAL-QUANTUM CHANNELS

6.1 Summary

Covert and secret quantum key distribution aims at generating information-theoretically secret bits between distant legitimate parties in a manner that remains provably undetectable by an adversary. We propose a framework in which to precisely define and analyze such an operation, and we show that covert and secret key expansion is possible. For fixed and known cq wiretap channels, we develop and analyze protocols based on forward and reverse reconciliation. The crux of our approach is the use of information reconciliation and privacy amplification techniques that are able to process the sparse signals required for covert operation and whose Shannon entropy scales as the square root of their length. In particular, our results show that the coordination required between legitimate parties to achieve covert communication can be achieved with a negligible number of secret key bits. The content of this chapter is based on [81].

6.2 Introduction

The possibility of quantum covert and secret key generation was recently explored [24, 82, 83] but has led to the rather pessimistic conclusion that "*covert QKD consumes more secret bits than it can generate*" [24]. Our main contribution is to offer a more nuanced and optimistic perspective and show that covert and secret key expansion is actually possible over quantum channels. The intuition behind our approach is the following. In layman's terms, the covertness constraint requires the number of qubit transmissions to scale as $O(\sqrt{n})$ for n channel uses [84]. A crucial characteristic of earlier works [84, 24] is that the scaling is ensured by having the legitimate parties *coordinate* the sparse transmission of $\sqrt{n}$ qubits in

channel uses chosen secretly and uniformly at random out of n. Unfortunately, the secret key size required to select these secret channel uses scales as $\Omega(\sqrt{n}\log n)$ and necessarily exceeds the number of covert bits that one can hope to obtain, which scales as $\Omega(\sqrt{n})$. In contrast, we introduce more sophisticated coding schemes for information reconciliation and privacy amplification that do not require such coordination and are able to directly process the sparse and diffuse statistical information content of covert signals. The protocols that we present do not yet offer the secrecy levels of state-of-the art QKD against coherent attacks but already achieve covert and secret key expansion over cq wiretap channels and might pave the way to more broadly applicable protocols.

Our results are developed in two steps as follows. We first lay out a precise model for quantum covert and secret key generation that captures a wide range of attacks by the adversary and protocols for legitimate parties, along with quantifiable metrics to assess the performance of a covert and secret key generation protocol over quantum channels. The main distinction with previous models [24, 82, 83] is the inclusion of the public communication required for information reconciliation in the analysis; specifically, since an adversary may devise a hypothesis test for detection based on all its observations, the probability distribution of the public communication has to be considered jointly with the quantum measurements in evaluating covertness. We then proceed to analyze an instance of quantum covert and secret key generation over fixed and known cq wiretap channels, for which we can define and analyze the covert and secret key capacity. We lower-bound the covert and secret key capacity by developing coding schemes using both forward and reverse reconciliation. The forward reconciliation scheme can be constructed by a suitable modification of established protocols for quantum covert communication [23] to guarantee secrecy. In contrast, the reverse reconciliation scheme requires a new approach because of technical challenges precluding the direct use of well-known results on information reconciliation and privacy amplification for the sparse distribution needed for covert communication. We do not instantiate explicit codes but recent progress in designing codes for

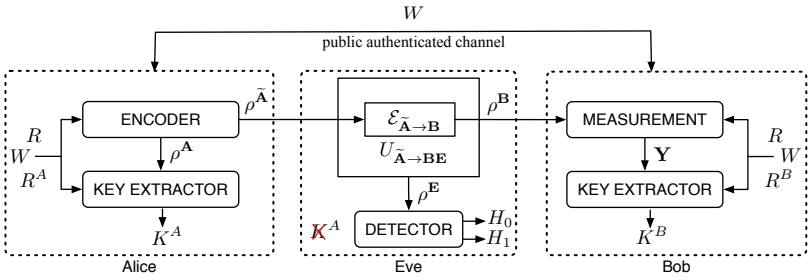

Figure 6.3.1: Model of covert and secret key expansion

covert communications [21] suggests that the protocols described here can be implemented with low-complexity.

6.3 Framework for Covert and Secret Key Generation over CQ Wiretap Channels

As illustrated in Fig. 6.3.1, we consider a setting in which two legitimate parties, Alice and Bob, desire to share a secret key while avoiding detection from an adversary, Eve, by exploiting a one-way quantum channel and a two-way classical authenticated public channel of unlimited capacity. Specifically, over n time steps, Alice prepares a cq state $\rho_{A\widetilde{A}}$, possibly depending on public communications, on a bipartite system described by a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_{\widetilde{A}}$ and sends the sub-system $\widetilde{A}$ to Bob. We assume that for $\mathcal{X} \subset \mathbb{R}$, $\{|x\rangle_A\}_{x\in\mathcal{X}}$ is an orthonormal basis for $\mathcal{H}_A$, all eigenvectors of ρ_A are always in $\{|x\rangle_A\}_{x\in\mathcal{X}}$, and for any $x \in \mathcal{X}$, the conditional state $\rho^x_{\widetilde{A}}$ is fixed. For simplicity, we restrict our attention to a two-dimensional $\mathcal{H}_A$, i.e., $\mathcal{X} = \{0, 1\}$, in which 0 represents an "innocent" symbol, corresponding to the absence of communication, while 1 represents an 'non-innocent" symbol. We further assume that the "start" and "stop" times of the protocol are known to all parties and obtained through other modalities, e.g., GPS signals. Eve expects the product state $(\rho^0_{\widetilde{A}})^{\otimes n}$ when there is no communication and may modify the states according to a quantum channel. We denote the entire state received by Bob and acting on the product Hilbert space $(\mathcal{H}_B)^{\otimes n}$ by $\rho_{\mathbf{B}}$.

For the purpose of covert communications, we need to distinguish protocols based on the type of Eve's attacks. In the most general case, Eve implements a *coherent attack*

described by a quantum channel

$$\mathcal{E}_{\widetilde{\mathbf{A}}\to\mathbf{B}} : \mathcal{L}\left(\mathcal{H}_{\widetilde{A}}^{\otimes n}\right) \to \mathcal{L}\left(\mathcal{H}_{B}^{\otimes n}\right), \tag{6.1}$$

with isomorphic extension $U_{\widetilde{\mathbf{A}}\to\mathbf{BE}}$, in which Bob receives $\rho_{\mathbf{B}} = \mathcal{E}_{\widetilde{\mathbf{A}}\to\mathbf{B}}(\rho_{\widetilde{\mathbf{A}}})$. In such situation, Bob should refrain from transmitting information over the public channel until the end of the transmission to avoid improving Eve's detection capability on the quantum channel. Note that this has no impact on QKD since no useful information is shared until the end of the protocol. A less powerful Eve can only implement *collective attacks* described by quantum channels of the form $\mathcal{E}_{\widetilde{\mathbf{A}}\to\mathbf{B}} = \mathcal{E}_{\widetilde{A}\to B}^{\otimes n}$, i.e., Eve applies the same channel independently to each state transmitted by Alice. In this case, we can assume that Bob receives each state before Alice transmits the next state, which allows meaningful public communication during the transmission between Alice and Bob. Throughout the paper, we assume that Alice and Bob have exact knowledge of the attack. We can therefore define an effective cq wiretap channel $x \mapsto \rho_{BE}^{x}$, with marginal cq channels $x \to \rho_{B}^{x}$ and $x \to \rho_{E}^{x}$ from Alice to Bob and Eve, respectively. Finally, Alice and Bob have access to independent local sources of randomness, denoted by $R^A \in \mathcal{R}^A$ and $R^B \in \mathcal{R}^B$, respectively, as well as a source of secret key $R \in \mathcal{R}$.

For simplicity, we describe the protocols with only reverse public communication, but extension to the general case, in which forward public communication is also allowed, does not present any difficulty. A protocol for key generation operates in n time steps as follows. Alice and Bob draw realizations r_A, r_B, and r of their local and common randomness. Subsequently, in every state $t \in [\![1, n]\!]$:

- Alice prepares a cq state $\rho_{A\widetilde{A}}$ as explained earlier using her local randomness r_A, the common randomness r, as well as past public messages from Bob denoted $(w_1, \cdots, w_{t-1})$ and sends $\rho_{\widetilde{A}}$ to Bob through the channel controlled by Eve;

- Bob performs a quantum measurement on his available quantum state to obtain a

classical measurement $y_t \in \mathcal{Y} \subset \mathbb{R}$;

- Bob sends a message $W_t \in \mathcal{W}_t$ over the public channel using his local randomness r_B, the common randomness r, as well as past measurements y^{t-1}. The choice of alphabet $\mathcal{W}_t$ is part of the protocol design.

At the end of time step n, *when no further public communication happens*, Eve performs a measurement on her state $\rho_{\mathbf{E}}$, as an attempt to detect the communication and obtain information about the secret key, while Alice and Bob use all their available information and randomness to compute two long binary strings s_X and s_Y, respectively, as well as the number of bits ℓ_X and ℓ_Y, respectively, to use as a secret key. The length of s_X and s_Y is public and fixed at the beginning of the protocol. Alice finally sets her key k_X to be the first ℓ_X bits of s_X while Bob sets his key k_Y to be the first ℓ_Y bits of s_Y.

A protocol is called an (ϵ, δ, μ)-protocol if the following properties hold. Let W, S_X, S_Y, K_X, K_Y, be the random variables representing the total public communication, Alice's random string, Bob's random string, Alice's key, and Bob's key, respectively. We require:

- ϵ-reliability: $P_e \triangleq \mathbb{P}(K_X \neq K_Y) \leqslant \epsilon$, which implicitly includes the condition $\ell_X = \ell_Y$;

- δ-secrecy: $S \triangleq \mathbb{D}\big(\rho_{\mathbf{E}WS_X} \| \rho_{\mathbf{E}W} \otimes \rho_{S_X}^{\text{unif}}\big) \leqslant \delta$, where $\rho_{\mathbf{E}WS_X}$ is the joint density matrix of the eavesdropper's observations, public messages and Alice's random string.

- μ-covertness: $C \triangleq \mathbb{D}\big(\rho_{\mathbf{E}W} \| \rho_{\mathbf{E}}^0 \otimes \rho_W^{\text{unif}}\big) \leqslant \mu$, where $\rho_{\mathbf{E}}^0$ is the density matrix of the eavesdropper's observations when no communication takes place.

A protocol is *efficient* if it allows key expansion so that the number of key bits created exceeds the number of common randomness bits consumed. Our goal is to analyze under what conditions efficient (ϵ, μ, δ)-protocols might exist.

A couple of remarks are in order regarding our protocol definition. Note that the choice of the key length is a part of the protocol. However, δ-secrecy requires the string S_X to be

secret and not just K_X. This is merely enforced for technical reasons, so that the relative entropy is a deterministic quantity irrespective of the length of the key. Since ϵ-reliability only applies to the bits of K_X, Alice can always generate the remaining bits of S_X independently and uniformly at random using her local randomness, so that our definition does not incur any loss of generality. By convention, we assume that the public communication is not by itself a proof of communication. Instead, μ-covertness only requires that the public bits look uniformly distributed and do not reveal communication on the quantum channel. We point out that δ-secrecy and μ-covertness are "one-shot" guarantees, in the sense that they only ensure a low probability of detection for a single execution of the protocol. In fact, by repeating the protocol k consecutive and independent times, a (ϵ, δ, μ)-protocol gives rise to a $(k\epsilon, k\delta, k\mu)$-protocol. Additional post-processing can reduce the constant $k\epsilon$ and $k\delta$ but cannot affect the constant $k\mu$. This suggests that the protocol should be designed for small values of μ and large values of n. Finally, the particular choice of the quantum state ρ_W^{unif} in the definition of covertness plays no role in our proofs. As long as there exists a specific state corresponding to no communication for the public communication, our proof holds and leads to a covert and secret key generation scheme.

6.4 Covert and Secret Key Generation over Known CQ Channels

We address the situation in which the cq wiretap channels are *fixed* and known ahead of time, and in which the adversary is *passive*. Our analysis corresponds to "known collective attacks." In this special case, the length of the key can be computed ahead of time, and there is no need to distinguish between the random strings S_X and S^Y and the keys K^X and K^Y. Furthermore, it becomes possible to define a notion of covert and secret key capacity as follows. A throughput Θ is achievable if there exists a sequence of $(\epsilon_n, \delta_n, \mu_n)$-protocols generating ℓ_n bits of secret key while consuming r_n bits of secret key over n stages and

such that

$$\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \delta_n = \lim_{n\to\infty} \mu_n = 0, \tag{6.2}$$

$$\ell_n = \omega(\log n), \tag{6.3}$$

$$\text{and } \lim_{n\to\infty} \frac{\ell_n - r_n}{\sqrt{n\mu_n}} \geqslant \Theta. \tag{6.4}$$

The supremum of all achievable throughputs is called the *covert and secret key capacity* and denoted C_{qck}. Note that the definition of the throughput already captures the scaling of the throughput with the square root of the number of channel uses, $\sqrt{n}$. The scaling is justified a posteriori by our analysis that shows that C_{qck} is lower bounded by a constant that only depends on the channel parameters. The unit of C_{qck} is therefore in nats per channel use. Our main results are lower-bounds on the covert capacity obtained by showing the existence of sequences of covert secret key generation protocols using reverse or forward reconciliation.

To analyze the performance of protocols with forward reconciliation, we build upon existing results for covert communication over cq channels [23, 42] with appropriate extensions to guarantee secrecy. The innovative principle of our approach is best highlighted for protocols with reverse reconciliation as follows. In a first phase, Alice transmits a sequence of iid symbols $\mathbf{X}$ distributed according to a Bernoulli(α_n) distribution over the cq channel, where $\alpha_n \in \omega((\frac{\log n}{n})^{\frac{2}{3}}) \cap o(\frac{1}{\sqrt{n}})$. Intuitively, the choice of $\{\alpha_n\}_{n\geqslant 1}$ must ensure that $\mathbf{X}$ is sparse, so that the warden cannot suspect the existence of information symbols, but not so sparse that Alice and Bob cannot extract a long enough key from their observation. We shall show that our choice of $\{\alpha_n\}_{n\geqslant 1}$ simultaneously satisfies both requirements. In a second phase, Bob measures his received quantum states in some basis and, based on the output of the measurements, generates two messages W and K, representing public information reconciliation and secret key, respectively. Bob subsequently sends W through the public channel, and Alice recovers K using W and $\mathbf{X}$. Although the second phase of

the protocol seems deceptively similar to a standard application of information reconciliation and privacy amplification, there exists a technical difficulty because of the specific distributions of Alice's and Bob's observations, which precludes the use of standard tools. Specifically, consider the classical channel $W_{Y|X}$ and suppose that $\mathbf{Y}$ is the output of the channel to the input $\mathbf{X}$. The standard finite-length analysis of reconciliation requires the second-order penalty γ_n to satisfy [35]

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n}\left(\log\frac{1}{W_{Y|X}(Y_i|X_i)} - \mathbb{H}(Y_i|X_i)\right) \geqslant \gamma_n\right) = 0. \tag{6.5}$$

By the Central Limit Theorem, this also requires that $\gamma_n^2 = \omega\left(\sum_{i=1}^{n} \mathrm{Var}\left(\frac{1}{W_{Y|X}(Y_i|X_i)}\right)\right)$. For our specific choice of α_n, one can check that $\mathrm{Var}\left(\frac{1}{W_{Y|X}(Y_i|X_i)}\right) = \Omega(1)$ so that the second-order penalty satisfies $\gamma_n = \omega(\sqrt{n})$. A similar reasoning holds for privacy amplification, which prevents us from establishing the desired first-order scaling of $o(\sqrt{n})$. We circumvent this difficulty by resorting to a technique called *likelihood encoder* [77], in which the encoders used to generate W and K are derived from different principles. In particular, the analysis of the likelihood encoder only requires the use of quantities depending on mutual information (instead of conditional entropy), which has the same scaling as the number of bits generated by a covert protocol. As we shall see later, instead of (6.5), the finite-length analysis of the likelihood encoder only requires the second-order penalty γ_n to satisfy

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n}\left(\log\frac{W_{Y|X}(Y_i|X_i)}{P_Y(Y_i)} - \mathbb{I}(X_i;Y_i)\right) \geqslant \gamma_n\right) = 0. \tag{6.6}$$

By the Central Limit Theorem, this now requires that $\gamma_n^2 = \omega\left(\sum_{i=1}^{n} \mathrm{Var}\left(\log\frac{W_{Y|X}(Y_i|X_i)}{P_Y(Y_i)}\right)\right)$. By our specific choice of α_n, one can check that $\mathrm{Var}\left(\log\frac{W_{Y|X}(Y_i|X_i)}{P_Y(Y_i)}\right) = O(\alpha_n)$, which leads to $\gamma_n = \omega(\sqrt{n\alpha_n})$. The second-order penalty is now conveniently dominated by the first-order term $\sum_{i=1}^{n}\mathbb{I}(X_i;Y_i)$ which is of the order of $\Omega(n\alpha_n)$.

The analysis of protocols with forward and reverse reconciliation leads to Theorem 26 below, whose proof is given in Appendix 6.A.

Theorem 26. *Let $\{|y\rangle_B\}$ be any orthonormal basis for $\mathcal{H}_B$, and define*

$$\widetilde{\rho}^x_{BE} \triangleq \sum_y \left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right) \rho^x_{BE} \left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right). \tag{6.7}$$

Assume that $\mathcal{H}_B$ and $\mathcal{H}_E$ have finite dimension and $0 < \chi_2\left(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\right) < \infty$.[1] *We have*

$$C_{\text{qck}} \geqslant \sqrt{\frac{2}{\chi_2\left(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\right)}} \left(\mathbb{D}\big(\widetilde{\rho}^1_B \| \widetilde{\rho}^0_B\big) - \mathbb{D}\big(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\big)\right), \tag{6.9}$$

and if $\widetilde{\rho}^0_{BE} = \widetilde{\rho}^0_B \otimes \widetilde{\rho}^0_E$, then

$$C_{\text{qck}} \geqslant \sqrt{\frac{2}{\chi_2\left(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\right)}} \left(\mathbb{D}\big(\widetilde{\rho}^1_{BE} \| \widetilde{\rho}^0_{BE}\big) - \mathbb{D}\big(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\big) - \mathbb{D}\big(\widetilde{\rho}^1_{BE} \| \widetilde{\rho}^1_B \otimes \widetilde{\rho}^1_E\big)\right), \tag{6.10}$$

which simplifies when $\widetilde{\rho}^1_{BE} = \widetilde{\rho}^1_B \otimes \widetilde{\rho}^1_E$ as

$$C_{\text{qck}} \geqslant \sqrt{\frac{2}{\chi_2\left(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\right)}} \mathbb{D}\big(\widetilde{\rho}^1_B \| \widetilde{\rho}^0_B\big). \tag{6.11}$$

In addition, the lower bound in (6.9) *is achieved without public communication using covert communication codes for cq channels [23] combined with wiretap coding techniques [13] while the lower bound in* (6.10) *is achieved with reverse reconciliation on the public channel.*

While this result does not hold for the most general quantum setting, note that the covert secret key throughputs predicted hold with a precise definition of covertness that

[1] χ_2 distance is

$$\chi_2\left(\rho \| \sigma\right) \triangleq \begin{cases} \operatorname{tr}\left(\rho^2 \sigma^{-1}\right) - 1 & \text{if } \operatorname{supp}(\rho) \subset \operatorname{supp}(\sigma), \\ \infty & \text{otherwise.} \end{cases} \tag{6.8}$$

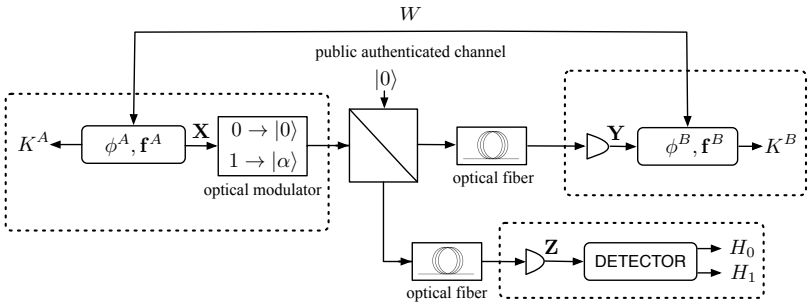

Figure 6.4.1: Simplified model of a lossy bosonic channel.

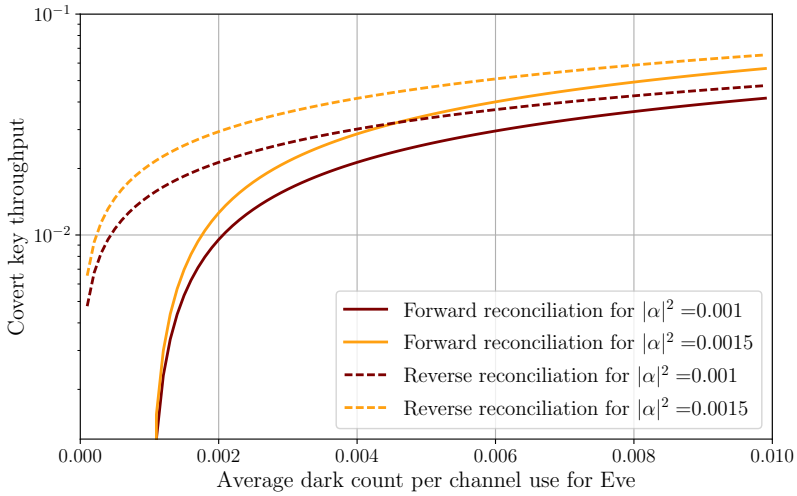

Figure 6.4.2: Covert and secret key generation throughput as a function of Eve's dark count rate.

explicitly includes the public communication and demonstrate the existence of efficient protocols that allow key expansion. Perhaps more importantly, as apparent in the proof of the result, such protocols do *not* rely on a secret key to determine the instances in which Alice transmits non-zero states; in contrast, our proof shows the existence of reconciliation and key-extraction algorithms capable of *extracting* the diffuse secret correlations created by Alice's sparse transmission of non-innocent states. We finally point out that for $\widetilde{\rho}^{\,1}_{BE} = \widetilde{\rho}^{\,1}_{B} \otimes \widetilde{\rho}^{\,1}_{E}$ secrecy comes *almost for free* as the information leakage to Eve is asymptotically dominated by the information shared between Alice and Bob in reverse reconciliation.

As an illustration, we consider the situation depicted in Fig. 6.4.1 in which the input port of a balanced beam-splitter is in control of Alice while Bob and Eve are each con-

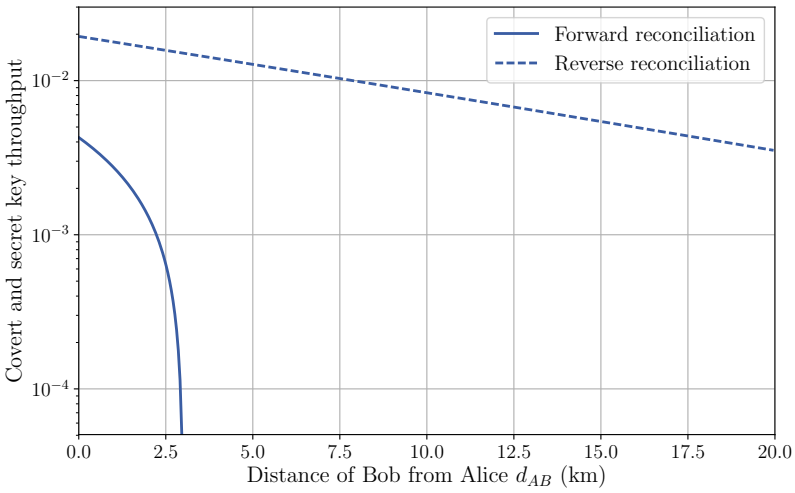

Figure 6.4.3: Covert and secret key generation throughput for a lossy bosonic channel.

nected to one of the output ports through optical fibers of length d_{AB} and d_{AE}, respectively, and loss γ dB/km. We further assume that the second input port is in the vacuum state, and that Alice uses the vacuum state $|0\rangle$ and a coherent state $|\alpha\rangle$ as the innocent and the information symbol, respectively. Bob and Eve measure their output ports with photodetectors to count the number of photons at each channel use. The photodetectors suffer from dark count that is beneficial for covert communication since detection of photons at Eve does not necessarily imply the existence communication. Let η_B and η_E be Bob and Eve's photodetector efficiency, respectively, and λ_B and λ_E be Bob and Eve's photodetector dark count rate, respectively. The achievable covert and secret key throughputs can be obtained by substituting the quantities

$$\widetilde{\eta}_B \triangleq \eta_B 10^{-\frac{d_{AB}\gamma}{10}} \tag{6.12}$$

$$\widetilde{\eta}_E \triangleq \eta_E 10^{-\frac{d_{AE}\gamma}{10}} \tag{6.13}$$

$$\chi_2\left(\widetilde{\rho}_E^1 \| \widetilde{\rho}_E^0\right) = e^{\frac{(\lambda_E+|\alpha|^2\widetilde{\eta}_E)^2}{\lambda_E} - \lambda_E + 2|\alpha|^2\widetilde{\eta}_E} \tag{6.14}$$

$$\mathbb{D}\left(\widetilde{\rho}_B^1 \| \widetilde{\rho}_B^0\right) = (\lambda_B + |\alpha|^2\widetilde{\eta}_B)\log\left(\lambda_B + |\alpha|^2\widetilde{\eta}_B\right) - |\alpha|^2\widetilde{\eta}_B \tag{6.15}$$

$$\mathbb{D}\left(\widetilde{\rho}_E^1 \| \widetilde{\rho}_E^0\right) = (\lambda_E + |\alpha|^2\widetilde{\eta}_E)\log\left(\lambda_E + |\alpha|^2\widetilde{\eta}_E\right) - |\alpha|^2\widetilde{\eta}_E \tag{6.16}$$

in (6.9) and (6.11) for forward and reverse reconciliation, respectively. Note that the output states of this channel belong to infinite-dimensional spaces and, strictly speaking, one cannot directly apply Theorem 26. Nevertheless, since for the number states $\{|n\rangle\}_{n\geqslant 0}$, $\langle n|\rho|n\rangle$ decays exponentially for all output states ρ, one can construct a sequence of channels with finite-dimensional output states for which the quantities used in (6.9) and (6.11), as well as the performance of any covert and secret key generation protocol, tend to those of the original channel.

We illustrate in Fig. 6.4.2 the achievable covert and secret key throughput as a function of Eve's photodetector dark count rate λ_E for $\gamma = 0.2$ dB/km, $\eta_B = \eta_E = 0.97$, $\lambda_B = 0.001$, and $d_{AB} = d_{AE} = 3$ km. In Fig. 6.4.3, we also illustrate the achievable covert and secret key throughput as a function of the distance of Bob to Alice d_{AB} for $|\alpha|^2 = 0.001$, $\gamma = 0.2$ dB/km, $\eta_B = \eta_E = 0.97$, $\lambda_B = \lambda_E = 0.001$, and $d_{AE} = 3$ km. As expected, the secret and covert key throughputs are orders of magnitude lower than their counterparts without covertness constraint. This is an unfortunate but unavoidable byproduct of the covertness constraint, which severely limits how many useful bits can be embedded in transmitted signals.

6.5 Conclusion

We have introduced a comprehensive framework in which to analyze the possibility of covert quantum key generation. In the special case of cq wiretap channels, for which the adversary's attack is known, we have established two lower-bounds on the optimal covert throughput of key generation based on forward and reverse reconciliation. While our results suggest that covert key expansion is possible over quantum channels, several lingering questions remain to be explored before envisioning an actual practical demonstration of covert quantum key distribution. This includes, in particular, extending our results to more general attacks with fewer assumptions regarding Eve's abilities, extending the analysis to infinite-dimensional systems that are closer to current technological implementations, and

designing efficient coding schemes with provable finite length performance. With respect to the latter, an explicit construction of covert communication codes over classical channels has been recently developed [21], which provides a promising lead to design codes for the framework proposed in the present paper.

APPENDIX

6.A Proof of Theorem 26

We prove Theorem 26 by generalizing the proof of [69, Theorem 1] to the quantum setting. The most challenging part of this generalization is to establish a channel resolvability result for cq channels for distributions suitable for covert communications. We first introduce some preliminary concepts regarding covert communications mostly borrowed from [13]. We also note that the use of standard proof techniques for secret key generation such as source coding with side information and privacy amplification is challenging for covert communication as discussed in Section 6.4. We therefore resort to the likelihood encoder technique [77] in which we first define an auxiliary problem that can be analyzed using channel coding approaches, for which designing a code for the main problem is reduced to the design of code for the auxiliary problem.

6.A.1 Preliminaries

We define here required quantities used for our achievability proof. Suppose Alice sends iid symbols through her cq channel $x \mapsto \rho_{BE}^x$ with each symbol distributed according to $Q_X \sim$ Bernoulli(α_n) for $\alpha_n \in (0,1)$. Upon receiving each state, Bob makes a measurement in a fixed orthonormal basis $\{|y\rangle_B\}$ for $\mathcal{H}_B$ to obtain a classical symbol y. In the following, we define equivalent cq channels from Bob to Alice and Eve that result in the same joint state for the three parties.

Definition 10. *Let* $\alpha_n \in [0,1]$. *We define*

$$Q_{Y|X}(y|x) \triangleq \langle y|_B \rho_B^x |y\rangle_B, \tag{6.17}$$

$$\widetilde{\rho}_{BE}^x \triangleq \sum_y \left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right) \rho_{BE}^x \left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right), \tag{6.18}$$

$$\widetilde{\rho}_{ABE} \triangleq \sum_x Q_X(x)|x\rangle\langle x|_A \otimes \widetilde{\rho}^x_{BE}, \tag{6.19}$$

$$\widetilde{\rho}^{x,y}_E \triangleq \frac{\mathrm{tr}_B\left(\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\rho^x_{BE}\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\right)}{Q_{Y|X}(y|x)} \tag{6.20}$$

$$\widetilde{\rho}^y_{AE} \triangleq \sum_x Q_{X|Y}(x|y)|x\rangle\langle x|_A \otimes \widetilde{\rho}^{x,y}_E. \tag{6.21}$$

Note that the state $\widetilde{\rho}_{ABE}$ is the joint state of all parties after Bob's measurement, which is classical for both Alice and Bob, and $\widetilde{\rho}^x_{BE}$, $\widetilde{\rho}^y_{AE}$, and $\widetilde{\rho}^{x,y}_E$ are the corresponding conditional quantum states.

The following lemma establishes useful properties of ρ^0_{BE} under the assumption $\widetilde{\rho}^0_{BE} = \widetilde{\rho}^0_B \otimes \widetilde{\rho}^0_E$.

Lemma 48. *If $\widetilde{\rho}^0_{BE} = \widetilde{\rho}^0_B \otimes \widetilde{\rho}^0_E$ then, for all y, it holds that $\widetilde{\rho}^{0,y}_E = \widetilde{\rho}^0_E$. Furthermore, we have*

$$I(Q_Y, \widetilde{\rho}^y_E) = \alpha_n\left(\mathbb{D}\big(\widetilde{\rho}^1_B \| \widetilde{\rho}^0_B\big) + \mathbb{D}\big(\widetilde{\rho}^1_E \| \widetilde{\rho}^0_E\big) - \right.$$
$$\left. \mathbb{D}\big(\widetilde{\rho}^1_{BE} \| \widetilde{\rho}^0_{BE}\big) + \mathbb{D}\big(\widetilde{\rho}^1_{BE} \| \widetilde{\rho}^1_B \otimes \widetilde{\rho}^1_E\big)\right) + O(\alpha_n^2). \tag{6.22}$$

Proof. By the spectral decomposition theorem, there exist orthonormal bases $\{|y\rangle_B\}$ and $\{|z\rangle_E\}$ for $\mathcal{H}_B$ and $\mathcal{H}_E$, respectively, such that

$$\widetilde{\rho}^0_B = \sum_y \lambda_y |y\rangle\langle y|_B \tag{6.23}$$

$$\widetilde{\rho}^0_E = \sum_z \lambda_z |z\rangle\langle z|_E \tag{6.24}$$

$$\widetilde{\rho}^0_{BE} = \sum_{y,y',z,z'} \lambda_{yy'zz'} |y\rangle\langle y'|_B \otimes |z\rangle\langle z'|_E. \tag{6.25}$$

Our assumption that $\widetilde{\rho}^0_{BE} = \widetilde{\rho}^0_B \otimes \widetilde{\rho}^0_E$ implies that $\lambda_{yy'zz'} = \lambda_y \lambda_z \mathbb{1}\{y = y', z = z'\}$. Fur-

thermore, for any y, we have by definition

$$\widetilde{\rho}_E^{0,y} \triangleq \frac{\mathrm{tr}_B\left(\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\rho_{BE}^0\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\right)}{Q_{Y|X}(y|0)} \tag{6.26}$$

$$= \frac{1}{Q_{Y|X}(y|0)}\mathrm{tr}_B\left(\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\left(\sum_{y',z'}\lambda_{y'}\lambda_{z'}|y'\rangle\langle y'|_B \otimes |z'\rangle\langle z'|_E\right)\right.$$

$$\left.\times\left(|y\rangle\langle y|_B \otimes \mathbf{1}_E\right)\right) \tag{6.27}$$

$$= \frac{\mathrm{tr}_B\left(\sum_{z'}\lambda_y\lambda_{z'}|y\rangle\langle y|_B \otimes |z'\rangle\langle z'|_E\right)}{Q_{Y|X}(y|0)} \tag{6.28}$$

$$= \frac{\lambda_y}{Q_{Y|X}(y|0)}\sum_{z'}\lambda_{z'}|z'\rangle\langle z'|_E \tag{6.29}$$

$$= \frac{\lambda_y}{Q_{Y|X}(y|0)}\widetilde{\rho}_E^0. \tag{6.30}$$

We also know that $\mathrm{tr}\left(\widetilde{\rho}_E^0\right) = \mathrm{tr}\left(\widetilde{\rho}_E^{0,y}\right) = 1$, which together with (6.30) yields $\widetilde{\rho}_E^0 = \widetilde{\rho}_E^{0,y}$.

To prove (6.22), notice that

$$I(Q_Y, \widetilde{\rho}_E^y) = \mathbb{I}(B;E)_{\widetilde{\rho}} \tag{6.31}$$

$$= \mathbb{I}(A;B)_{\widetilde{\rho}} + \mathbb{I}(A;E)_{\widetilde{\rho}} - \mathbb{I}(A;BE)_{\widetilde{\rho}} + \mathbb{I}(B;E|A)_{\widetilde{\rho}}. \tag{6.32}$$

Moreover, for $\widetilde{\rho}_B^{\alpha_n} \triangleq (1-\alpha_n)\widetilde{\rho}_B^0 + \alpha_n\widetilde{\rho}_B^1$, we can write

$$\mathbb{I}(A;B)_{\widetilde{\rho}} = H(\widetilde{\rho}_B^{\alpha_n}) - (1-\alpha_n)H(\widetilde{\rho}_B^0) - \alpha_n H(\widetilde{\rho}_B^1) \tag{6.33}$$

$$= -\mathrm{tr}(\widetilde{\rho}_B^{\alpha_n}\log\left(\widetilde{\rho}_B^{\alpha_n}\right) - (1-\alpha_n)\widetilde{\rho}_B^0\log(\widetilde{\rho}_B^0) - \alpha_n\widetilde{\rho}_B^1\log(\widetilde{\rho}_B^1)) \tag{6.34}$$

$$= -\mathrm{tr}(\widetilde{\rho}_B^{\alpha_n}\left(\log\left(\widetilde{\rho}_B^{\alpha_n}\right) - \log\left(\widetilde{\rho}_B^0\right) + \log\left(\widetilde{\rho}_B^0\right)\right)$$

$$- (1-\alpha_n)\widetilde{\rho}_B^0\log(\widetilde{\rho}_B^0) - \alpha_n\widetilde{\rho}_B^1\log(\widetilde{\rho}_B^1)) \tag{6.35}$$

$$= -\mathrm{tr}(\widetilde{\rho}_B^{\alpha_n}\left(\log\widetilde{\rho}_B^{\alpha_n} - \log\widetilde{\rho}_B^0\right) - \alpha_n\widetilde{\rho}_B^1\left(\log\widetilde{\rho}_B^1 - \log\widetilde{\rho}_B^0\right)) \tag{6.36}$$

$$= \alpha_n\mathbb{D}\big(\widetilde{\rho}_B^1\|\widetilde{\rho}_B^0\big) - \mathbb{D}\big(\widetilde{\rho}_B^{\alpha_n}\|\widetilde{\rho}_B^0\big) \tag{6.37}$$

$$\overset{(a)}{=} \alpha_n\mathbb{D}\big(\widetilde{\rho}_B^1\|\widetilde{\rho}_B^0\big) + O(\alpha_n^2), \tag{6.38}$$

where (a) follows from [23, Equation (19)]. Similarly, we obtain

$$\mathbb{I}(A;E)_{\widetilde{\rho}} = \alpha_n \mathbb{D}\big(\widetilde{\rho}_E^1 \| \widetilde{\rho}_E^0\big) + O(\alpha_n^2), \tag{6.39}$$

$$\mathbb{I}(A;BE)_{\widetilde{\rho}} = \alpha_n \mathbb{D}\big(\widetilde{\rho}_{BE}^1 \| \widetilde{\rho}_{BE}^0\big) + O(\alpha_n^2). \tag{6.40}$$

Since X is classical, [85, Equation (11.92)] yields that

$$\mathbb{I}(B;E|A)_{\widetilde{\rho}} = (1-\alpha_n)\mathbb{I}(B;E)_{\widetilde{\rho}^0} + \alpha_n \mathbb{I}(B;E)_{\widetilde{\rho}^1} \tag{6.41}$$

$$\overset{(a)}{=} \alpha_n \mathbb{I}(B;E)_{\widetilde{\rho}^1} \tag{6.42}$$

$$= \alpha_n \mathbb{D}\big(\widetilde{\rho}_{BE}^1 \| \widetilde{\rho}_B^1 \otimes \widetilde{\rho}_E^1\big), \tag{6.43}$$

where (a) follows from our assumption that $\widetilde{\rho}_{BE}^0 = \widetilde{\rho}_B^0 \otimes \widetilde{\rho}_E^0$. This completes the proof of (6.22). □

6.A.2 An Auxiliary Problem

To show the existence of good codes for our main problem, we use the likelihood encoder technique [77], and in particular, define an auxiliary problem for which we can exploit channel coding instead of source coding. We then show how these two problems are related in Section 6.A.3. Consider a cq channel $y \mapsto \widetilde{\rho}_{AE}^y$ from Bob to Alice and Eve as in Definition 10. Bob encodes three uniformly distributed messages $W_1 \in [\![1, M_1]\!]$, $W_2 \in [\![1, M_2]\!]$, and $W_3 \in [\![1, M_3]\!]$ into a codeword $\mathbf{Y}$ using an encoder $f : [\![1, M_1]\!] \times [\![1, M_2]\!] \times [\![1, M_3]\!] \to \mathcal{Y}^n$, transmits the codeword $\mathbf{Y}$ over the cq channel, and sends W_2 publicly. Alice subsequently performs a measurement on her received state $\rho_{\mathbf{Y}}^a$ in a fixed basis $\{|x\rangle\}$ to obtain $\mathbf{X}$, and uses $\mathbf{X}$ and W_2 to decode W_1 as $\widehat{W_1}$. If $P_{\mathbf{Y}}^a$ denotes the induced PMF of $\mathbf{Y}$, and $\rho^a_{\mathbf{ABE}W_1W_2W_3\widehat{W_1}}$ is the joint state in the auxiliary problem, our objective is to ensure that $\mathbb{P}\Big(\widehat{W_1} \neq W_1\Big)$, $\frac{1}{2}\|P_{\mathbf{Y}}^a - Q_Y^{\otimes n}\|_1$, and $\|\rho_{\mathbf{E}W_1W_2} - \rho_{\mathbf{E}} \otimes \rho_{W_1W_2}\|_1$ are small.

Lemma 49. *If for some $\zeta > 0$*

$$\log M_3 = \lfloor (1+\zeta) I(Q_Y, \widetilde{\rho}_E^y) n \rfloor, \tag{6.44}$$

$$\log M_1 + \log M_2 + \log M_3 = \lceil (1+\zeta) H(Q_Y) n \rceil, \tag{6.45}$$

$$\log M_1 + \log M_3 = \lfloor (1-\zeta) I(Q_Y, Q_{X|Y}) n \rfloor, \tag{6.46}$$

then there exists a sequence of codes and a positive constant ξ such that

$$\mathbb{P}\Big(\widehat{W}_1 \neq W_1\Big) \leqslant 2^{-\xi \alpha_n n}, \tag{6.47}$$

$$\frac{1}{2} \| P_{\mathbf{Y}}^a - Q_Y^{\otimes n} \|_1 \leqslant 2^{-\xi n}, \tag{6.48}$$

$$\| \rho_{\mathbf{E} W_1 W_2} - \rho_{\mathbf{E}} \otimes \rho_{W_1 W_2} \|_1 \leqslant 2^{-\omega(\log n)}. \tag{6.49}$$

Proof. Let $F : [\![1, M_1]\!] \times [\![1, M_2]\!] \times [\![1, M_3]\!]$ be a random encoder whose codewords are drawn independently according to $Q_Y^{\otimes n}$. By construction, Alice can assume that each symbol X_i is received as the output of a DMC $(\mathcal{Y}, Q_{X|Y}, \mathcal{X})$ with input Y_i, and, therefore, Lemma 3 in Chapter 1 implies that

$$\mathbb{E}_F\Big(\mathbb{P}\Big(\widehat{W}_1 \neq W_1\Big)\Big) = \frac{1}{M_2} \sum_{w_2} \mathbb{E}_F\Big(\mathbb{P}\Big(\widehat{W}_1 \neq W_1 | W_2 = w_2\Big)\Big) \tag{6.50}$$

$$\overset{(a)}{\leqslant} \mathbb{P}_{Q_{X|Y}^{\otimes n} \otimes Q_Y^{\otimes n}}\left(\sum_{t=1}^n \log \frac{Q_{X|Y}(X_t|Y_t)}{Q_X(X_t)} \leqslant \gamma \right) + \frac{M_1 M_3}{2^\gamma} \tag{6.51}$$

$$= \mathbb{P}_{Q_{XY}^{\otimes n}}\left(\sum_{t=1}^n \log \frac{Q_{Y|X}(Y_t|X_t)}{Q_Y(Y_t)} \leqslant \gamma \right) + \frac{M_1 M_3}{2^\gamma}, \tag{6.52}$$

where (a) follows from applying Lemma 3 to the subcodebook $\{F(w_1, w_2, w_3) : w_1 \in [\![1, M_1]\!], w_3 \in [\![1, M_3]\!]\}$ for a particular w_2. By choosing

$$\log M_1 + \log M_3 = \lfloor (1-\zeta) I(Q_X, Q_{Y|X}) n \rfloor \tag{6.53}$$

$$\gamma = \left(1 - \frac{\zeta}{2}\right) I(Q_X, Q_{Y|X}) n, \tag{6.54}$$

and using Bernstein's inequality (Theorem 4 in Chapter 1), we obtain

$$\begin{aligned}\mathbb{P}_{Q_{XY}^{\otimes n}}\left(\sum_{t=1}^{n}\log\frac{Q_{Y|X}(Y_t|X_t)}{Q_Y(Y_t)}\leqslant\gamma\right)+\frac{M_1M_3}{2^{\gamma}}&\\ \leqslant\exp\left(-\frac{-\frac{1}{8}\zeta^2 I(Q_Y,Q_{X|Y})^2 n}{\mathrm{Var}\left(\log\frac{Q_{Y|X}(Y|X)}{Q_Y(Y)}\right)+\frac{1}{3}C_3\zeta\mathbb{I}(X;Y)}\right)+2^{-\frac{\zeta}{2}\mathbb{I}(X;Y)n}&\\ \leqslant 2^{-\xi\alpha_n n},&\end{aligned}\tag{6.55}$$

for some $\xi>0$. Next, by using Lemma 5 in Chapter 1 for the channel $(\mathcal{Y},Q_{Y'|Y},\mathcal{Y})$ with $Q_{Y'|Y}(y'|y)\triangleq\mathbb{1}\{y'=y\}$ and the distribution Q_Y, we obtain

$$\mathbb{E}_F\left(\frac{1}{2}\|P^a_{\mathbf{Y}};Q_Y^{\otimes n}\|_1\right)\leqslant\mathbb{P}_{Q_Y^{\otimes n}}\left(\sum_{t=1}^{n}\log\frac{1}{Q_Y(Y_t)}\geqslant\gamma\right)+\sqrt{\frac{2^{\gamma}}{M_1M_2M_3}}.\tag{6.56}$$

By choosing

$$\log M_1+\log M_2+\log M_3=\lceil(1+\zeta)\mathbb{H}(Y)\,n\rceil\tag{6.57}$$

$$\gamma=\left(1+\frac{\zeta}{2}\right)\mathbb{H}(Y)\,n\tag{6.58}$$

and using Hoeffding's inequality (Theorem 3 in Chapter 1), with $\mu_Y\triangleq\min_{y:Q_Y(y)>0}Q_Y(y)$, we obtain

$$\mathbb{P}_{Q_Y^{\otimes n}}\left(\sum_{t=1}^{n}\log\frac{1}{Q_Y(Y_t)}\geqslant\gamma\right)+\sqrt{\frac{2^{\gamma}}{M_1M_2M_3}}\leqslant\exp\left(-\frac{\zeta^2\mathbb{H}(Y)^2\,n}{2\log^2(\mu_Y)}\right)+2^{-\frac{\zeta}{2}\mathbb{H}(Y)n}\tag{6.59}$$

$$\leqslant 2^{-\xi n},\tag{6.60}$$

for $\xi>0$ small enough.

Since W_1 and W_2 are classical, we can write

$$\rho_{W_1W_2\mathbf{E}} = \frac{1}{M_1M_2}\sum_{w_1,w_2} |w_1w_2\rangle\langle w_1w_2| \otimes \rho_{\mathbf{E}}^{w_1w_2}. \tag{6.61}$$

To upper-bound $\mathbb{E}_F(\|\rho_{\mathbf{E}W_1W_2} - \rho_{\mathbf{E}} \otimes \rho_{W_1W_2}\|_1)$, we apply Lemma 6 in Chapter 1 and obtain

$$\mathbb{E}_F\big(\|\rho_{W_1W_2\mathbf{E}} - \rho_{W_1W_2} \otimes (\widetilde{\rho}_E)^{\otimes n}\|_1\big) = \frac{1}{M_1M_2}\sum_{w_1,w_2} \mathbb{E}_F\big(\|\rho_{\mathbf{E}}^{w_1,w_2} - (\widetilde{\rho}_E)^{\otimes n}\|_1\big) \tag{6.62}$$

$$\leqslant \sqrt{2^{\gamma s + n\phi(s)}} + \sqrt{\frac{2^{\gamma}\nu}{M_3}}, \tag{6.63}$$

where ν is the number of distinct eigenvalues of $(\widetilde{\rho}_E)^{\otimes n}$, and

$$\phi(s) = \log\left(\sum_y Q_Y(y)\text{tr}\left((\widetilde{\rho}_E^y)^{1-s}(\widetilde{\rho}_E)^s\right)\right). \tag{6.64}$$

Upon choosing

$$\log M_3 = \lfloor I(Q_Y, \widetilde{\rho}_E^y)n + \zeta\alpha_n n\rfloor, \tag{6.65}$$

$$\gamma = I(Q_Y, \widetilde{\rho}_E^y)n + \frac{\zeta}{2}\alpha_n n, \tag{6.66}$$

we obtain

$$\sqrt{2^{\gamma s + n\phi(s)}} + \sqrt{\frac{2^{\gamma}\nu}{M_3}} \leqslant \sqrt{2^{s\alpha_n n\left(\frac{I(Q_Y,\widetilde{\rho}_E^y)}{\alpha_n} + \frac{\zeta}{2} + \frac{\phi(s)}{s\alpha_n}\right)}} + \sqrt{2^{-\frac{\zeta}{2}\alpha_n n}\nu} \tag{6.67}$$

$$\overset{(a)}{\leqslant} \sqrt{2^{s\alpha_n n\left(\frac{I(Q_Y,\widetilde{\rho}_E^y)}{\alpha_n} + \frac{\zeta}{2} + \frac{\phi(s)}{s\alpha_n}\right)}} + \sqrt{2^{-\frac{\zeta}{2}\alpha_n n}(n+1)^{\dim\mathcal{H}_E}} \tag{6.68}$$

$$\leqslant \sqrt{2^{s\alpha_n n\left(\frac{I(Q_Y,\widetilde{\rho}_E^y)}{\alpha_n} + \frac{\zeta}{2} + \frac{\phi(s)}{s\alpha_n}\right)}} + \frac{1}{2}2^{-\xi\alpha_n n}, \tag{6.69}$$

where (a) follows from [37, Lemma 3.7]. We now introduce the following technical lemma

to simplify the above expression.

Lemma 50. *Suppose $s < 0$; there exists a constant $B \geqslant 0$ such that for n large enough and $|s|$ small enough, we have*

$$\phi(s) > -I(Q_Y, \widetilde{\rho}_E^y)s - B(\alpha_n s^2 - s^3). \tag{6.70}$$

Proof. See Appendix 6.B. □

Applying Lemma 50 to (6.69), we obtain

$$\sqrt{2^{s\alpha_n n\left(\frac{I(Q_Y,\widetilde{\rho}_E^y)}{\alpha_n}+\frac{\zeta}{2}+\frac{\phi(s)}{s\alpha_n}\right)}} \leqslant \sqrt{2^{s\alpha_n n\left(\frac{I(Q_Y,\widetilde{\rho}_E^y)}{\alpha_n}+\frac{\zeta}{2}+\frac{-I(Q_Y,\widetilde{\rho}_E^y)s-B(\alpha_n s^2-s^3)}{s\alpha_n}\right)}} \tag{6.71}$$

$$= \sqrt{2^{s\alpha_n n\left(\frac{\zeta}{2}+\frac{B(\alpha_n s-s^2)}{\alpha_n}\right)}}. \tag{6.72}$$

By choosing $s = o(\sqrt{\alpha_n}) \cap \omega(\frac{\log n}{n\alpha_n})$[2], the above expression goes to zero faster than any polynomial. Therefore, for a random encoder, we have

$$\mathbb{E}_F\Big(\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big)\Big) \leqslant 2^{-\xi\alpha_n n} \tag{6.73}$$

$$\mathbb{E}_F\left(\frac{1}{2}\|P^a_{\mathbf{Y}} - Q_Y^{\otimes n}\|_1\right) \leqslant 2^{-\xi n} \tag{6.74}$$

$$\mathbb{E}_F\big(\|\rho_{W_1W_2\mathbf{E}} - (\widetilde{\rho}_E)^{\otimes n} \otimes \rho_{W_1W_2}\|_1\big) \leqslant 2^{-\omega(\log n)}, \tag{6.75}$$

if

$$\log M_3 = \lfloor (1+\zeta) I(Q_Y, \widetilde{\rho}_E^y) n \rfloor, \tag{6.76}$$

$$\log M_1 + \log M_2 + \log M_3 = \lceil (1+\zeta) H(Q_Y) n \rceil, \tag{6.77}$$

$$\log M_1 + \log M_3 = \lfloor (1-\zeta) I(Q_Y, Q_{X|Y}) n \rfloor. \tag{6.78}$$

[2]To find such s, it is required that $\sqrt{\alpha_n} = \omega(\frac{\log n}{n\alpha_n})$ or equivalently $\alpha_n = \omega\left(\left(\frac{\log n}{n}\right)^{\frac{2}{3}}\right)$

Upon defining the events

$$\mathcal{E}_1 \triangleq \{\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big) \leqslant 4 \times 2^{-\xi\alpha_n n}\}, \tag{6.79}$$

$$\mathcal{E}_2 \triangleq \{\frac{1}{2}\|P^a_{\mathbf{Y}} - Q_Y^{\otimes n}\|_1 \leqslant 4 \times 2^{-\xi n}\}, \tag{6.80}$$

$$\mathcal{E}_3 \triangleq \{\|\rho_{W_1W_2\mathbf{E}} - (\widetilde{\rho}_E)^{\otimes n} \otimes \rho_{W_1W_2}\|_1 \leqslant 4 \times 2^{-\omega(\log n)}\}, \tag{6.81}$$

and using Markov inequality, we have

$$\mathbb{P}_F(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) \geqslant 1 - \mathbb{P}_F(\mathcal{E}_1^c) - \mathbb{P}_F(\mathcal{E}_2^c) - \mathbb{P}_F(\mathcal{E}_3^c) \tag{6.82}$$

$$\geqslant 1 - \frac{\mathbb{E}_F\Big(\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big)\Big)}{2^{-\xi\alpha_n n}} - \frac{\mathbb{E}_F\big(\frac{1}{2}\|P^a_{\mathbf{Y}} - Q_Y^{\otimes n}\|_1\big)}{4 \times 2^{-\xi n}} \tag{6.83}$$

$$- \frac{\mathbb{E}_F\big(\|\rho_{W_1W_2\mathbf{E}} - (\widetilde{\rho}_E)^{\otimes n} \otimes \rho_{W_1W_2}\|_1\big)}{4 \times 2^{-\omega(\log n)}} \tag{6.84}$$

$$\geqslant \frac{1}{4}. \tag{6.85}$$

Therefore, there exists a realization f of F with

$$\mathbb{P}\Big(W_1 \neq \widehat{W}_1\Big) \leqslant 4 \times 2^{-\xi\alpha_n n}, \tag{6.86}$$

$$\frac{1}{2}\|P^a_{\mathbf{Y}} - Q_Y^{\otimes n}\|_1 \leqslant 4 \times 2^{-\xi n}, \tag{6.87}$$

$$\|\rho_{W_1W_2\mathbf{E}} - (\widetilde{\rho}_E)^{\otimes n} \otimes \rho_{W_1W_2}\|_1 \leqslant 4 \times 2^{-\omega(\log n)}. \tag{6.88}$$

□

6.A.3 Proof of Theorem 26

Using the likelihood encoder technique, we first prove the lower bound in (6.10). Consider a specific code for the auxiliary problem in Section 6.A.2 and let $\widetilde{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}$ be the corresponding induced joint quantum state. Because all random variables W_1, W_2, $\mathbf{X}$, and $\mathbf{Y}$ are classical, we can define their induced joint PMF denoted by $\widetilde{P}_{W_1W_2\mathbf{XY}}$. We then use

the conditional PMFs $\widetilde{P}_{W_1W_2|\mathbf{Y}}$ and $\widetilde{P}_{\widehat{W}_1|\mathbf{X}W_2}$ as the encoder and decoder, respectively, in the main problem, which results in the induced joint quantum state $\widehat{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}$. By our construction, we can decompose $\widetilde{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}$ as

$$\widetilde{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1} = \sum_{w_1,w_2,\widehat{w}_1,\mathbf{y},\mathbf{x}} \widetilde{P}_{\mathbf{Y}}(\mathbf{y})\widetilde{P}_{W_1W_2|\mathbf{Y}}(w_1,w_2|\mathbf{y})Q_{X|Y}^{\otimes n}(\mathbf{x}|\mathbf{y})\widetilde{P}_{\widehat{W}_1|\mathbf{X}W_2}(\widehat{w}_1|\mathbf{x},w_2) \times |\mathbf{y}\mathbf{x}w_1w_2\widehat{w}_1\rangle\langle\mathbf{y}\mathbf{x}w_1w_2\widehat{w}_1| \otimes \widetilde{\rho}_{\mathbf{E}}^{\mathbf{x},\mathbf{y}}, \quad (6.89)$$

and $\widehat{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}$ as

$$\widehat{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1} = \sum_{w_1,w_2,\widehat{w}_1,\mathbf{y},\mathbf{x}} Q_Y^{\otimes n}(\mathbf{y})\widetilde{P}_{W_1W_2|\mathbf{Y}}(w_1,w_2|\mathbf{y})Q_{X|Y}^{\otimes n}(\mathbf{x}|\mathbf{y})\widetilde{P}_{\widehat{W}_1|\mathbf{X}W_2}(\widehat{w}_1|\mathbf{x},w_2) \times |\mathbf{y}\mathbf{x}w_1w_2\widehat{w}_1\rangle\langle\mathbf{y}\mathbf{x}w_1w_2\widehat{w}_1| \otimes \widetilde{\rho}_{\mathbf{E}}^{\mathbf{x},\mathbf{y}}. \quad (6.90)$$

Since they differ only in the distribution of $\mathbf{Y}$, we have

$$\|\widetilde{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1} - \widehat{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}\|_1 \leqslant \left\|\widetilde{P}_{\mathbf{Y}}^a - Q_Y^{\otimes n}\right\|_1 \overset{(a)}{\leqslant} 2^{-\xi n}, \quad (6.91)$$

where (a) follows from (6.74). Thus, we upper-bound the probability of error in the main problem as

$$\begin{aligned}\mathbb{P}_{\widehat{P}}(W_1 \neq W_2) &\leqslant \mathbb{P}_{\widetilde{P}}(W_1 \neq W_2) + \|\widetilde{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1} - \widehat{\rho}_{\mathbf{ABE}W_1W_2\widehat{W}_1}\|_1 &(6.92)\\ &\leqslant 2^{-\zeta\alpha_n n} + 2^{-\zeta n}, &(6.93)\end{aligned}$$

and upper-bound the sum of secrecy and covertness as

$$\begin{aligned}S + C &\triangleq \mathbb{D}\big(\widehat{\rho}_{W_1W_2\mathbf{E}}\|\rho_{W_1}^{\text{unif}} \otimes \widehat{\rho}_{W_2\mathbf{E}}\big) + \mathbb{D}\big(\widehat{\rho}_{W_2\mathbf{E}}\|\rho_{W_2}^{\text{unif}} \otimes \big(\rho_E^0\big)^{\otimes n}\big) &(6.94)\\ &= \mathbb{D}\big(\widehat{\rho}_{W_1W_2\mathbf{E}}\|\rho_{W_1W_2}^{\text{unif}} \otimes \widehat{\rho}_{\mathbf{E}}\big) + \mathbb{D}\big(\widehat{\rho}_{\mathbf{E}}\|\big(\rho_E^0\big)^{\otimes n}\big) &(6.95)\\ &\overset{(a)}{=} \mathbb{D}\big(\widehat{\rho}_{W_1W_2\mathbf{E}}\|\rho_{W_1W_2}^{\text{unif}} \otimes \widehat{\rho}_{\mathbf{E}}\big) + \frac{1}{2}\alpha_n^2\chi_2(\rho_E^1\|\rho_E^0)n + O(\alpha_n^3 n) &(6.96)\end{aligned}$$

$$\overset{(b)}{\leqslant} \|\widehat{\rho}^{W_1W_2\mathbf{E}} - \rho_{\text{unif}}^{W_1W_2} \otimes \widehat{\rho}^{\mathbf{E}}\|_1 \log \frac{M_1M_2\left(\dim \mathcal{H}_E\right)^n}{\frac{1}{M_1M_2}\lambda_{min}(\widetilde{\rho}_E)^n \|\widehat{\rho}^{W_1W_2\mathbf{E}} - \rho_{\text{unif}}^{W_1W_2} \otimes \widehat{\rho}^{\mathbf{E}}\|_1} + \frac{1}{2}\alpha_n^2 \chi_2(\rho_E^1\|\rho_E^0)n + O(\alpha_n^3 n) \tag{6.97}$$

$$= \|\widehat{\rho}^{W_1W_2\mathbf{E}} - \rho_{\text{unif}}^{W_1W_2} \otimes \widehat{\rho}^{\mathbf{E}}\|_1 \left(O(n) - \log \|\widehat{\rho}^{W_1W_2\mathbf{E}} - \rho_{\text{unif}}^{W_1W_2} \otimes \widehat{\rho}^{\mathbf{E}}\|_1\right) + \frac{1}{2}\alpha_n^2 \chi_2(\rho_E^1\|\rho_E^0)n + O(\alpha_n^3 n) \tag{6.98}$$

$$\overset{(c)}{\leqslant} \left(2^{-\zeta\alpha_n n} + 2^{-\zeta n}\right) O(n) + \frac{1}{2}\alpha_n^2 \chi_2(\rho_E^1\|\rho_E^0)n + O(\alpha_n^3 n), \tag{6.99}$$

where (a) follows from [23, Lemma 7], (b) follows from Lemma 51 in Appendix 6.C, and (c) follows from

$$\|\widehat{\rho}_{W_1W_2\mathbf{E}} - \rho_{W_1W_2}^{\text{unif}} \otimes \widehat{\rho}_{\mathbf{E}}\|_1 \leqslant \|\widetilde{\rho}_{W_1W_2\mathbf{E}} - \rho_{W_1W_2}^{\text{unif}} \otimes \widehat{\rho}_{\mathbf{E}}\|_1 + \|\widehat{\rho}_{W_1W_2\mathbf{E}} - \widetilde{\rho}_{W_1W_2\mathbf{E}}\|_1 \tag{6.100}$$

$$\leqslant 2^{-\zeta\alpha_n n} + 2^{-\zeta n}. \tag{6.101}$$

The throughput of the coding scheme is lower-bounded by (6.104) shown below.

$$\frac{\log M_1}{\sqrt{nC}} \geqslant \frac{\log M_1}{\sqrt{n\left(\left(2^{-\zeta\alpha_n n} + 2^{-\zeta n}\right) O(n) + \frac{1}{2}\alpha_n^2 \chi_2(\rho_E^1\|\rho_E^0)n + O(\alpha_n^3 n)\right)}} \tag{6.102}$$

$$\geqslant \sqrt{\frac{2}{\chi_2(\rho_E^1\|\rho_E^0)}} \frac{\lfloor (1-\zeta)\mathbb{I}(A;B)_{\widetilde{\rho}} n \rfloor - \lceil \mathbb{I}(B;E)_{\widetilde{\rho}} n + \zeta\alpha_n n \rceil}{n\alpha_n(1+o(1))} \tag{6.103}$$

$$= \sqrt{\frac{2}{\chi_2\left(\widetilde{\rho}_E^1\|\widetilde{\rho}_E^0\right)}} \left(\mathbb{D}\big(\widetilde{\rho}_{BE}^1\|\widetilde{\rho}_{BE}^0\big) - \mathbb{D}\big(\widetilde{\rho}_E^1\|\widetilde{\rho}_E^0\big) - \mathbb{D}\big(\widetilde{\rho}_{BE}^1\|\widetilde{\rho}_B^1 \otimes \widetilde{\rho}_E^1\big)\right) + o(1). \tag{6.104}$$

We finally turn to the proof of the lower bound in (6.9). Note that if $\mathbb{D}(\widetilde{\rho}_B^1\|\widetilde{\rho}_B^0) \leqslant \mathbb{D}(\widetilde{\rho}_E^1\|\widetilde{\rho}_E^0)$, the result is trivial. Therefore, we can assume that $\mathbb{D}(\widetilde{\rho}_B^1\|\widetilde{\rho}_B^0) > \mathbb{D}(\widetilde{\rho}_E^1\|\widetilde{\rho}_E^0)$.

Let M_1 and M_2 be such that

$$\log M_1 + \log M_2 = \lfloor (1-\zeta) I(Q_X, \widetilde{\rho}_B^y) \rfloor, \tag{6.105}$$

$$\log M_2 = \lceil (1+\zeta) I(Q_X, \widetilde{\rho}_E^y) \rceil. \tag{6.106}$$

The protocol is then as follows. Alice chooses a random binary string of length $\log M_1 + \log M_2$ and transmits this string through a covert code introduced in [23]. Alice and Bob subsequently extract the first $\log M_1$ bits of the string as the key. The reliability and covertness proof follows exactly from [23]. For secrecy, note that

$$\mathbb{D}\big(\rho_{\mathbf{E}MS^X} \| \rho_{\mathbf{E}M} \otimes \rho_{S^X}^{\text{unif}}\big) \triangleq \mathbb{D}\big(\rho_{\mathbf{E}S^X} \| \rho_{\mathbf{E}} \otimes \rho_{S^X}^{\text{unif}}\big) \tag{6.107}$$

$$= \frac{1}{M_1} \sum_{w_1=1}^{M_1} \mathbb{D}(\rho_{\mathbf{E}}^{w_1} \| \rho_{\mathbf{E}}). \tag{6.108}$$

Similar to the proof of (6.75), one can show that the above expression is upper-bounded by $2^{-\omega(\log n)}$ provided that $\log M_2 = \lceil (1+\zeta) I(Q_X, \widetilde{\rho}_E^y) \rceil$. Lower-bounding the throughput as in (6.104) using (6.105) and (6.106) concludes the proof. Note that unlike the proof of (6.10), the protocol used here does not use the public communication channel.

6.B Proof of Lemma 50

For a fix n, applying Taylor's theorem on ϕ defined in (6.64), we have

$$\phi(s) = \phi(0) + \phi'(0)s + \frac{\phi''(0)}{2}s^2 + \frac{\phi'''(\eta)}{6}s^3, \tag{6.109}$$

for some $s \leqslant \eta \leqslant 0$. To compute derivatives of ϕ, let us define

$$A_y(s) \triangleq (\widetilde{\rho}_E^y)^{1-s} (\widetilde{\rho}_E)^s, \tag{6.110}$$

$$g(s) \triangleq \sum_y Q_Y(y) \text{tr}\left(A_y(s)\right). \tag{6.111}$$

One can check that $\phi(s) = \log g(s)$. Hence, we obtain

$$\phi'(s) = \frac{g'(s)}{g(s)}, \tag{6.112}$$

$$\phi''(s) = \frac{g''(s)}{g(s)} - \left(\frac{g'(s)}{g(s)}\right)^2, \tag{6.113}$$

$$\phi'''(s) = \frac{g'''(s)}{g(s)} - 3\frac{g'(s)g''(s)}{g^2(s)} + 2\left(\frac{g'(s)}{g(s)}\right)^3. \tag{6.114}$$

Moreover, since $A'_y(s) = -\ln\left(\widetilde{\rho}_E^y\right) A_y(s) + A_y(s)\ln\left(\widetilde{\rho}_E\right)$, we have

$$g'(s) = \sum_y Q_Y(y)\mathrm{tr}\left(-\ln\left(\widetilde{\rho}_E^y\right) A_y(s) + A_y(s)\ln\left(\widetilde{\rho}_E\right)\right), \tag{6.115}$$

$$g''(s) = \sum_y Q_Y(y)\mathrm{tr}\Big(\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^2 A_y(s) \\ -2\ln\left(\widetilde{\rho}_E^y\right) A_y(s)\ln\left(\widetilde{\rho}_E\right) + A_y(s)\left(\ln\left(\widetilde{\rho}_E\right)\right)^2\Big), \tag{6.116}$$

and

$$g'''(s) = \sum_y Q_Y(y)\mathrm{tr}\Big(-\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^3 A_y(s) + 3\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^2 A_y(s)\ln\left(\widetilde{\rho}_E\right) \\ -3\ln\left(\widetilde{\rho}_E^y\right) A_y(s)\left(\ln\left(\widetilde{\rho}_E\right)\right)^2 + A_y(s)\left(\ln\left(\widetilde{\rho}_E\right)\right)^3\Big).$$

Using $A_y(0) = \widetilde{\rho}_E^y$ combined with the above expressions, we obtain

$$g(0) = \sum_y Q_Y(y) \mathrm{tr}\left(\widetilde{\rho}_E^y\right) = 1, \tag{6.117}$$

$$g'(0) = \sum_y Q_Y(y) \mathrm{tr}\left(-\ln\left(\widetilde{\rho}_E^y\right) \widetilde{\rho}_E^y + \widetilde{\rho}_E^y \ln\left(\widetilde{\rho}_E\right)\right) \tag{6.118}$$

$$= -I(Q_Y, \widetilde{\rho}_E^y), \tag{6.119}$$

$$g''(0) = \sum_y Q_Y(y) \mathrm{tr}\left(\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^2 \widetilde{\rho}_E^y - 2\ln\left(\widetilde{\rho}_E^y\right) \widetilde{\rho}_E^y \ln\left(\widetilde{\rho}_E\right) + \widetilde{\rho}_E^y \left(\ln\left(\widetilde{\rho}_E\right)\right)^2\right). \tag{6.120}$$

Hence, we have

$$\phi(0) = \ln(g(0)) = 0, \tag{6.121}$$

$$\phi'(0) = \frac{g'(0)}{g(0)} = -I(Q_Y, \widetilde{\rho}_E^y), \tag{6.122}$$

$$\phi''(0) = \frac{g''(0)}{g(0)} - \left(\frac{g'(0)}{g(0)}\right)^2, \tag{6.123}$$

$$= \sum_y Q_Y(y) \mathrm{tr}\left(\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^2 \widetilde{\rho}_E^y - 2\ln\left(\widetilde{\rho}_E^y\right) \widetilde{\rho}_E^y \ln\left(\widetilde{\rho}_E\right) + \widetilde{\rho}_E^y \left(\ln\left(\widetilde{\rho}_E\right)\right)^2\right)$$

$$- I(Q_Y, \widetilde{\rho}_E^y)^2. \tag{6.124}$$

Note that $\phi''(0)$ implicitly depends on α_n the probability that the input is one. Let us define

$$h(\alpha) \triangleq \sum_y Q_Y(y) \mathrm{tr}\left(\left(\ln\left(\widetilde{\rho}_E^y\right)\right)^2 \widetilde{\rho}_E^y - 2\ln\left(\widetilde{\rho}_E^y\right) \widetilde{\rho}_E^y \ln\left(\widetilde{\rho}_E\right) + \widetilde{\rho}_E^y \left(\ln\left(\widetilde{\rho}_E\right)\right)^2\right) \tag{6.125}$$

when the input distribution is Bernoulli(α). One can check that $Q_Y(y)$, $\widetilde{\rho}_E^y$, $\ln(\widetilde{\rho}_E^y)$, and $\ln(\widetilde{\rho}_E)$ are continuously differentiable with respect to α, and so is h. Moreover, we have

$$h(0) = \sum_y Q_{Y|X}(y|0)\text{tr}\left(\left(\ln\left(\widetilde{\rho}_E^{0,y}\right)\right)^2 \widetilde{\rho}_E^{0,y} - 2\ln\left(\widetilde{\rho}_E^{0,y}\right)\widetilde{\rho}_E^{0,y}\ln\left(\widetilde{\rho}_E\right) + \widetilde{\rho}_E^{0,y}\left(\ln\left(\widetilde{\rho}_E\right)\right)^2\right) \tag{6.126}$$

$$\overset{(a)}{=} \sum_y Q_{Y|X}(y|0)\text{tr}\left(\left(\ln\left(\widetilde{\rho}_E\right)\right)^2 \widetilde{\rho}_E - 2\ln\left(\widetilde{\rho}_E\right)\widetilde{\rho}_E\ln\left(\widetilde{\rho}_E\right) + \widetilde{\rho}_E\left(\ln\left(\widetilde{\rho}_E\right)\right)^2\right) = 0, \tag{6.127}$$

where (a) follows from Lemma 48. By the mean value theorem, we know that $|h(\alpha) - h(0)| = |h(\alpha)| = h'(\beta)\alpha$ for some $0 < \beta < \alpha$. Since h' is continuous for a small neighborhood around zero, it is bounded and therefore, we have $|h(\alpha_n)| = O(\alpha_n)$. Furthermore, Lemma 48 implies that $I(Q_Y, \widetilde{\rho}_E^y)^2 = O(\alpha_n^2)$. Thus, there exists $B > 0$ such that $|\phi''(0)| \leqslant B\alpha_n$ for n large enough. Notice next that g, g', g'', and g''' are jointly continuous functions of both variables s and α_n in a neighborhood around (0, 0). Additionally, since $g(0) = 1$ when $\alpha = 0$, we conclude that ϕ''' is also continuous in both s and α_n in a neighborhood around (0, 0). Therefore, for B large enough, $|s|$ small enough and n large enough, we have $|\phi'''(s)| \leqslant B$. Combing $\phi(0) = 0$, $\phi'(0) = -I(Q_Y, \widetilde{\rho}_E^y)$, $|\phi''(0)| \leqslant B\alpha_n$, and $|\phi'''(\eta)| \leqslant B$ with (6.109), we obtain the desired result.

6.C Technical Lemma

Lemma 51. *Suppose ρ and σ are two density matrices on a Hilbert space $\mathcal{H}$ with* $\dim \mathcal{H} = d$ *such that* $\text{supp}\rho \subset \text{supp}\sigma$ *and* $\|\rho - \sigma\|_1 \leqslant \epsilon \leqslant e^{-1}$. *Then,*

$$\mathbb{D}(\rho\|\sigma) \leqslant \epsilon \log \frac{d}{\lambda_{\min}(\sigma)\epsilon}. \tag{6.128}$$

Proof. Since $\text{supp}(\rho) \subset \text{supp}(\sigma)$, we have

$$\mathbb{D}(\rho\|\sigma) = \text{tr}\left(\rho(\log\rho - \log\sigma)\right) \tag{6.129}$$

$$= -H(\rho) + H(\sigma) - \text{tr}\left((\rho-\sigma)\log\sigma\right) \tag{6.130}$$

$$\overset{(a)}{\leqslant} \epsilon \log\frac{d}{\epsilon} - \text{tr}\left((\rho-\sigma)\log\sigma\right) \tag{6.131}$$

$$\leqslant \epsilon\log\frac{d}{\epsilon} + \epsilon\log\frac{1}{\lambda_{\min}(\sigma)}, \tag{6.132}$$

where (a) follows from Fannes inequality. □

CHAPTER 7
COVERT SECRET KEY GENERATION: TOWARD A PARCTICAL EXPERIMENT OVER A QUANTUM CHANNEL

7.1 Summary

We propose a protocol based on pulse-position modulation and multi-level coding that allows one to bootstrap traditional quantum key distribution protocols while ensuring covertness, in the sense that no statistical test by the adversary can detect the presence of communication over the quantum channel better than a random guess. When run over a bosonic channel, our protocol can leverage existing discrete-modulated continuous variable protocols. Since existing techniques to bound Eve's information do not directly apply, we develop a new bound that results in positive, although very low, throughput for a range of channel parameters. The analysis of the protocol performance shows that covert secret key expansion is possible using a public authenticated classical channel and a quantum channel largely but not fully under control of an adversary, which we precisely define. We also establish a converse result showing that, under the golden standard of quantum key distribution, by which the adversary completely controls the quantum channel, no covert key generation is possible. The content of this chapter is based on [86].

7.2 Introduction

The first attempts at covert QKD [24, 84] have ensured covertness with *fully coordinated* protocols, in which information-bearing qubits are only communicated over a secret random subset of channel uses upon which Alice and Bob secretly agree prior to communication; in the remaining channel uses Alice transmits an "idle" state corresponding to no communication. If n denotes the total number of channel uses and t denotes the number

of channel uses over which communication happens, fully coordinated protocols [24, 84] require $t = \Theta(\sqrt{n})$ to generate $\Omega(\sqrt{n})$ bits of secret key. Although the processing complexity is identical to that of standard QKD protocols, fully coordinated protocols require Alice and Bob to share $\log \binom{n}{t} = \Theta(\sqrt{n} \log n)$ secret bits prior to communication, so that the number of required key bit asymptotically dominates the number of generated key bits, thereby forbidding secret key expansion.

To circumvent the impossibility of covert and secret key expansion with fully coordinated protocols, we have proposed in Chapter 6 to achieve covertness with an *uncoordinated* protocol based on the use of "sparse signaling" for quantum state distribution, which operates as follows. If $\alpha_n \triangleq O(n^{-\frac{1}{2}})$ and if P_X denotes the Bernoulli(α_n) distribution, Alice generates an i.i.d. sequence $X^n = (X_1, \cdots, X_n)$ according to $P_X^{\otimes n}$, which is then modulated by mapping zero to the idle state and one to another state. A technical subtlety, however, prevents Alice and Bob from performing classical information reconciliation and privacy amplification to obtain a secret key from their shared quantum states. While the asymptotic key rate is $O(n^{-\frac{1}{2}})$ by the square root law, the finite length penalty of privacy amplification is of the order of $\omega(n^{-\frac{1}{2}})$ [87], which dominates the asymptotic rate. For a *known adversary's attack*, our uncoordinated protocol circumvents this difficulty and ensures secret key expansion using a likelihood encoder but the classical post-processing of the protocol *is much more complex than for typical QKD protocols*.

To reap the benefits of both fully coordinated and uncoordinated protocols and achieve secret key expansion without increasing processing complexity, we develop here a *partially coordinated* protocol inspired by our prior construction of low-complexity codes for covert communication over classical channels with PPM and MLC [21]. This approach is more aligned with traditional low-complexity information reconciliation and privacy amplification algorithms and we analyze the covertness and the security under an unknown attack by the adversary. We restrict, however, the adversary's attack by requiring that a portion of the channel be out of the adversary's control (e.g., the part of channel in Alice's laboratory).

We prove that such a requirement is fundamentally necessary to establish any covertness result. Since we were not able to use any standard technique to bound Eve's information, we present a new bound, which we use to show the existence of positive throughputs for some range of bosonic channel parameters. While our results are slightly disappointing in that the range of useful parameters is limited, our analysis opens the way to experimental demonstrations of covert QKD.

7.3 Notation

In this chapter, we used the following two notations. We define $C(\rho_A, \sigma_A) \triangleq \sqrt{1 - F(\rho_A, \sigma_A)}$, which satisfies the triangle inequality [88, Proposition 3.3]. For a non-empty finite set $\mathcal{X}$, let $\mathcal{H}_X$ be a Hilbert space defined by an orthonormal basis $\{|x\rangle : x \in \mathcal{X}\}$. For a function $f : \mathcal{X} \to \mathcal{Y}$, we define the channel

$$\begin{aligned} \mathcal{E}^f_{X\to Y} : \mathcal{L}(\mathcal{H}_X) &\to \mathcal{L}(\mathcal{H}_Y) \\ \rho_X &\mapsto \sum_{x\in\mathcal{X}} |f(x)\rangle\langle x|\rho_X|x\rangle\langle f(x)|. \end{aligned} \tag{7.1}$$

7.4 Covert QKD Setup

Alice and Bob aim at *covertly* expanding a *secret* key using the following generic setup and protocol. Let R_A and R_B be Alice's and Bob's local randomness, respectively, and let R be a secret common randomness. As depicted in Figure 7.3.1, Alice has a transmitter in her

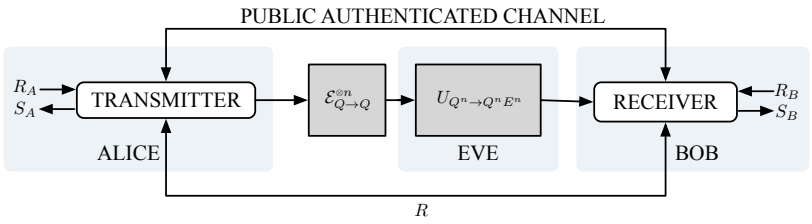

Figure 7.3.1: Covert quantum key expansion model in the presence of Eve

laboratory to send quantum states to Bob. At any time instant, the state of the transmitter is described by a density operator on a Hilbert space $\mathcal{H}_Q$. A pure state $|0\rangle\langle 0|$ identifies the "idle" state of the transmitter when there is no communication.[1] Alice prepares a quantum state $\widetilde{\sigma}_{AQ^n} = \mathrm{tr}_{RR_AR_B}\left(\widetilde{\sigma}_{RR_AR_BAQ^n}\right)$ and sends $\widetilde{\sigma}_{Q^n}$ to Bob by n uses of her transmitter. The adversary Eve is assumed to receive the state through a known memoryless quantum channel, which we call *probe*, $\mathcal{E}_{Q\to Q}$ that is *outside its control*. Eve therefore obtains the output of $\mathcal{E}_{Q\to Q}^{\otimes n}$ for the input $\widetilde{\sigma}_{Q^n}$, which then interacts with an ancilla E^n in Eve's lab before being transmitted to Bob. The whole operation can be described by an isometry $U_{Q^n\to Q^nE^n}$, for which we denote the corresponding quantum channel by $\mathcal{U}_{Q^n\to Q^nE^n}$. We call this phase *quantum state distribution*, which results in the joint quantum state

$$\sigma_{RR_AR_BAQ^nE^n} \triangleq (\mathrm{id}_{RR_AR_BA} \otimes \mathcal{U}_{Q^n\to Q^nE^n} \circ \mathcal{E}_{Q\to Q}^{\otimes n})(\widetilde{\sigma}_{RR_AR_BAQ^n}) \tag{7.2}$$

between Alice, Bob, and Eve, respectively. After establishing a shared quantum state, Alice and Bob may interactively communicate over an authenticated classical public channel and perform measurements on their available state to generate keys S_A and S_B, respectively. We call this phase *quantum key distillation* and formally describe it by a quantum channel $\mathcal{D}_{R_AR_BRAQ^n\to CS_AS_B}$, where C denotes all public communication. The final state is then

$$\sigma_{CS_AS_BE^n} \triangleq (\mathrm{id}_{E^n} \otimes \mathcal{D}_{R_AR_BRAQ^n\to CS_AS_B})(\sigma_{RR_AR_BAQ^nE^n}). \tag{7.3}$$

[1] One can associate a mixed state to no communication, but in bosonic systems, the natural choice for the idle state is a pure vacuum state.

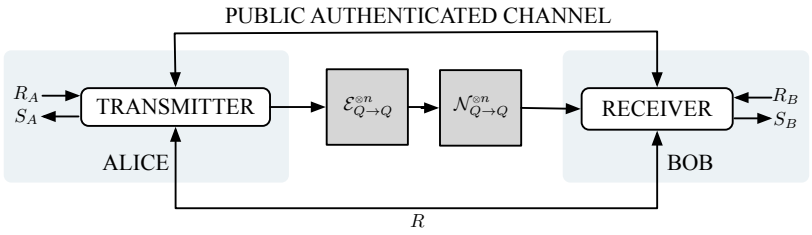

Figure 7.3.2: Covert quantum key expansion model in the absence of Eve

Finally, we assume that, in the absence of an adversary, Alice and Bob expect to be connected through the "honest" channel $\mathcal{N}_{Q\to Q}$ *after the probe* (see Fig 7.3.2). Alice and Bob can also abort the protocol at any time and do not generate secret keys.

For a particular protocol inducing the final joint state $\sigma_{CS_AS_BE^n}$, we assess the performance of the protocol with the following four quantities:

1. the probability of error $\mathbb{P}(S_A \neq S_B|\text{not abort})$;
2. the information leakage $\left\|\sigma_{S_AE^nC} - \rho^{\text{unif}}_{S_A} \otimes \sigma_{E^nC}\right\|_1$;
3. the covertness $\left\|\sigma_{CE^n} - \rho^{\text{unif}}_{C} \otimes \rho^0_{E^n}\right\|_1$, where $\rho^0_{E^n} \triangleq \mathcal{U}_{Q^n\to E^n}(\mathcal{E}^{\otimes n}_{Q\to Q}(|0\rangle\langle 0|^{\otimes n}))$; and
4. the robustness $\mathbb{P}(\text{abort})$ in the presence of the honest channel $\mathcal{N}_{Q\to Q}$.

We highlight here three crucial distinctions between our model and traditional QKD.

1. As covertness is of no concern in QKD, the idle state of the transmitter is not specified in a QKD model.
2. Unlike QKD, in which the quantum channel is in complete control of the adversary, we restrict Eve's observations to result from a known probe $\mathcal{E}_{Q\to Q}$. We discuss this limitation of our model in Section 7.5.
3. To the best of our knowledge, there is no standard way of defining covertness in the presence of public communication in the literature. We use a covertness criterion similar to that in Chapter 6, in which the mere existence of public communication does not reveal the existence of the protocol; however, when $\left\|\sigma_{CE^n} - \rho^{\text{unif}}_{C} \otimes \rho^0_{E^n}\right\|_1$ is small, we effectively require negligible dependence between public communication and σ_{E^n} and that public communication be distributed according to a prespecified distribution, which we choose to be the uniform distribution ρ^{unif}_C for simplicity. These two requirements are critical to ensure that public communication does not help Eve detect the communication over the quantum channel.

7.5 Role of the Probe

We now establish a no-go result in the absence of the probe. We measure here the information leakage and the covertness through the relaxed matrics $\left\|\sigma_{S_A C} - \rho_{S_A}^{\text{unif}} \otimes \sigma_C\right\|_1$ and $\|\widetilde{\sigma}_{Q^n} - \rho_{E^n}^0\|_1$, respectively, instead of $\left\|\sigma_{S_A E^n C} - \rho_{S_A}^{\text{unif}} \otimes \sigma_{E^n C}\right\|_1$ and $\left\|\sigma_{C E^n} - \rho_C^{\text{unif}} \otimes \rho_{E^n}^0\right\|_1$. Note that

$$\left\|\sigma_{S_A C} - \rho_{S_A}^{\text{unif}} \otimes \sigma_C\right\|_1 \leqslant \left\|\sigma_{S_A E^n C} - \rho_{S_A}^{\text{unif}} \otimes \sigma_{E^n C}\right\|_1$$

and

$$\left\|\widetilde{\sigma}_{Q^n} - \rho_{E^n}^0\right\|_1 \leqslant \left\|\sigma_{C E^n} - \rho_C^{\text{unif}} \otimes \rho_{E^n}^0\right\|_1$$

by the data processing inequality. Thus, a converse for the relaxed secrecy and covertness constraints implies a converse for the constraint as defined in Section 7.4.

Theorem 27. *Let $\mathcal{E}_{Q\to Q} = \text{id}_Q$ and define $K \triangleq \log \dim \mathcal{H}_{S_A}$. Consider a protocol that operates as in Section 7.4 with $\mathbb{P}(S_A \neq S_B) \leqslant \epsilon$, $\left\|\sigma_{S_A C} - \rho_{S_A}^{\text{unif}} \otimes \sigma_C\right\|_1 \leqslant \delta$, and $\left\|\widetilde{\sigma}_{Q^n} - |0\rangle\langle 0|^{\otimes n}\right\|_1 \leqslant \mu$. We then have*

$$(1 - 5\sqrt{\mu} - \epsilon - 2\delta)K \leqslant \mathbb{H}_b\left(\sqrt{\mu}\right) + \mathbb{H}_b\left(\epsilon + \sqrt{\mu}\right) + 2\left(1 + \sqrt{\mu}\right)\mathbb{H}_b\left(\frac{\sqrt{\mu}}{1 + \sqrt{\mu}}\right). \quad (7.4)$$

Proof. See Appendix 7.A. □

Consequently, if $\epsilon, \delta, \mu \to 0$ then K vanishes, as well. Theorem 27 therefore shows that giving the *complete* control of the channel to the adversary is too stringent to establish covertness. A probe is therefore necessary and could be created with some part of the channel that is protected from the adversary, for example the portion of an optical fiber that lies inside Alice's laboratory.

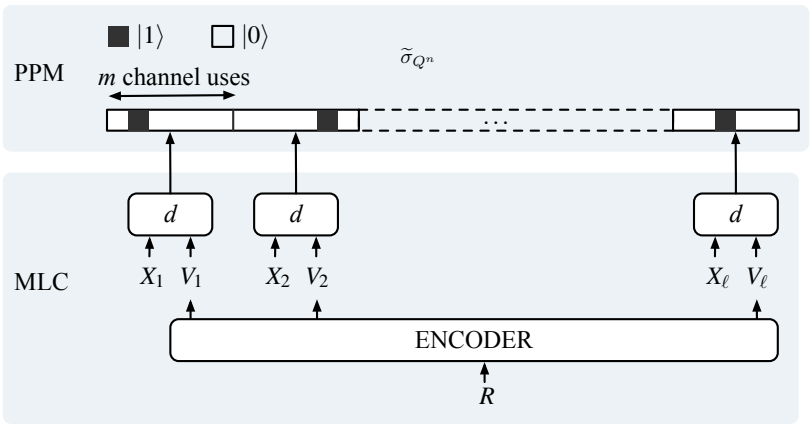

Figure 7.6.1: Covert quantum state distribution through PPM and MLC

7.6 Description of PPM-MLC-based Protocol

We first provide a high level description of the role of PPM and MLC in our PPM-MLC-based protocol. The principle of PPM is to split the whole transmission block into smaller sub-blocks and to transmit exactly one non-idle state in a position chosen uniformly at random in each sub-block. The number of sub-blocks and the size of each sub-block should both be $O(\sqrt{n})$ to achieve covertness [46]. The idea behind MLC is to further split the randomness used to specify the position of the non-idle state into two parts: one part with a fixed size independent of n, generated locally by Alice and used for key generation, and another part of size growing with n, generated secretly and jointly by Alice and Bob and used for mimicking the uniform distribution via quantum channel resolvability [37, Chapter 9.4]. This splitting allows Alice and Bob to *partially* coordinate without paying the penalty incurred by full coordination. The use of MLC converts the problem of covert QKD into a traditional QKD problem over an effective block-length scaling as $O(\sqrt{n})$, for which low-complexity processing is possible.

We now elaborate on the details of the partially coordinated PPM-MLC-based protocol. As depicted in Fig. 7.6.1, the n channel uses are partitioned into ℓ consecutive sub-blocks

of length m so that $n \triangleq \ell m$. Fix a non-idle state $|\phi\rangle$ for the transmitter such that

$$\langle\phi|0\rangle \neq 0, \tag{7.5}$$

$$\operatorname{supp}\left(\mathcal{E}_{Q\to Q}(|1\rangle\langle 1|)\right) \subseteq \operatorname{supp}\left(\mathcal{E}_{Q\to Q}(|0\rangle\langle 0|)\right). \tag{7.6}$$

We define the z^{th} PPM state of length m, $|\mathrm{PPM}, z\rangle_{Q^m}$ as

$$|0\rangle^{\otimes z-1} \otimes |\phi\rangle \otimes |0\rangle^{\otimes m-z}, \tag{7.7}$$

a product state of $|0\rangle$ and $|\phi\rangle$ with a single non-idle state in the z^{th} position. Writing $m \triangleq m_x m_v$, Alice generates ℓ PPM states of length m by choosing the position of the non-idle state in the i^{th} state as

$$d(X_i, V_i) \triangleq (X_i - 1)m_v + V_i, \tag{7.8}$$

where $X^\ell = (X_1, \cdots, X_\ell) \in [\![1, m_x]\!]^\ell$ and $V^\ell = (V_1, \cdots, V_\ell) \in [\![1, m_v]\!]^\ell$ are randomly generated sequences. Let $\rho_{Q^n}^{x^\ell, v^\ell}$ be the corresponding density operator when $X^\ell = x^\ell$ and $V^\ell = v^\ell$, i.e.,

$$\rho_{Q^n}^{x^\ell, v^\ell} \triangleq \otimes_{i=1}^{\ell} |\mathrm{PPM}, d(x_i, v_i)\rangle\langle \mathrm{PPM}, d(x_i, v_i)| \tag{7.9}$$

The crux of the PPM-MLC-based protocol is to generate the sequences X^ℓ and V^ℓ using *different mechanisms:* X^ℓ is generated locally by Alice i.i.d. according to the uniform distribution over $[\![1, m_x]\!]$ while V^ℓ is generated jointly by Alice and Bob by sampling codewords uniformly at random from a codebook of size h described as follows. Let $\mathcal{F}$ be a regular two-universal family of hash functions from $[\![1, m_v]\!]^\ell \to \mathcal{Z}$ where $\mathcal{Z} = [\![1, \frac{{m_v}^\ell}{h}]\!]$. Bob samples $F \in \mathcal{F}$ and $Z \in \mathcal{Z}$ uniformly at random and transmits them over the public channel. The codebook consists of the codewords in $F^{-1}(Z)$ and will be denoted through

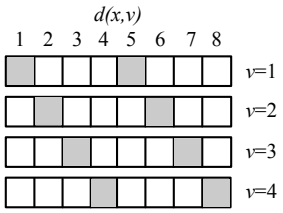

Figure 7.6.2: Gray squares indicate the possible values of $d(x, v)$ for a given value of v when $m_x = 2$ and $m_v = 4$.

the function

$$g : [\![1, h]\!] \to [\![1, m_v]\!]^\ell : R \mapsto V^\ell = g(R). \tag{7.10}$$

The choice of X^ℓ uniformly at random defines an effective cq channel from v^ℓ to the state at the output of the probe, formally described by

$$v^\ell \mapsto \frac{1}{{m_x}^\ell} \sum_{x^\ell} \mathcal{E}_{Q\to Q}^{\otimes n} \left(\rho_{Q^n}^{x^\ell, v^\ell} \right). \tag{7.11}$$

By sampling R uniformly at random in $[\![1, h]\!]$ and using $g(R)$ at the input of the effective cq channel, Eve's received state is

$$\sigma_{Q^n} \triangleq \frac{1}{{m_x}^\ell} \frac{1}{h} \sum_{x^\ell, r} \mathcal{E}_{Q\to Q}^{\otimes n} \left(\rho_{Q^n}^{x^\ell, g(r)} \right). \tag{7.12}$$

If Alice and Bob secretly share R prior to the transmission, Bob can discard $m - m_x$ of his sub-systems in each sub-block of length m, for which he knows that the state $|0\rangle$ is sent (see Fig. 7.6.2 for an example). We shall later account for the partial coordination through R by subtracting $\log h$ from the number of generated key bits. For each sub-block, Alice therefore obtains the classical state X_i while Bob obtains m_x received states. We denote the whole state shared between Alice and Bob in ℓ sub-blocks by $\sigma_{X^\ell (Q^{m_x})^\ell}$, which is $\tau_{XQ^{m_x}}^{\otimes \ell}$ in the absence of the adversary for some $\tau_{XQ^{m_x}}$ independent of n.

The last steps of the PPM-MLC-based protocol are similar to a traditional QKD proto-

col applied to $\sigma_{X^{\ell}(Q^{m_x})^{\ell}}$ with the additional constraint $\left\|\sigma_{CE^n} - \rho_C^{\text{unif}} \otimes \rho_{E^n}^0\right\|_1 \leqslant \delta$, which requires the public communication to be uniformly distributed and independent of Eve's observation during the quantum communication phase. The three main steps of this phase are parameter estimation, information reconciliation, and privacy amplification. Let $\ell \triangleq \ell_1 + \ell_2$ and decompose $X^{\ell}(Q^{m_x})^{\ell}$ into two disjoint parts $X^{\ell_1}(Q^{m_x})^{\ell_1}$ and $X^{\ell_2}(Q^{m_x})^{\ell_2}$, used for parameter estimation and secret key distillation, respectively. For simplicity, we do not detail the classical algorithm for information reconciliation and take for granted the existence of a protocol $\mathcal{I}_{X^{\ell_2}(Q^{m_x})^{\ell_2} \to X^{\ell_2}\widehat{X}^{\ell_2}C_{\text{IR}}}$ where $\widehat{X}^{\ell_2}$ denotes Bob's estimate of X^{ℓ_2} and C_{IR} is the public communication that takes place during the information reconciliation protocol. Let $\sigma_{X^{\ell_2},\widehat{X}^{\ell_2}C_{IR}E^nC'} \triangleq (\mathcal{I} \otimes \text{id}_{E^nC'})(\sigma_{X^{\ell_2}(Q^{m_x})^{\ell_2}E^nC'})$ where σ_{E^n} is the adversary's observation from the quantum communication and C' is the public communication in the quantum state distribution phase. We assume that there exists a known ϵ_{IR} such that

$$\mathbb{P}\left(X^{\ell_2} \neq \widehat{X}^{\ell_2}\right) \leqslant \epsilon_{\text{IR}}, \tag{7.13}$$

$$\mathbb{P}(\text{abort}|\text{ honest channel}) \leqslant \epsilon_{\text{IR}}, \tag{7.14}$$

$$\left\|\sigma_{C_{\text{IR}}E^nC'} - \rho_{C_{\text{IR}}}^{\text{unif}} \otimes \sigma_{E^nC'}\right\|_1 \leqslant \epsilon_{\text{IR}}. \tag{7.15}$$

More details to justify the existence of good reconciliation protocols can be found in [89] and references therein. Furthermore, it is shown in [90] that the public communication could be uniformly distributed and independent of the adversary's observations as required in (7.15) by using a negligible amount of key. The final step is to perform privacy amplification to establish a secure key. To this end, Alice and Bob require a bound on $\mathbb{H}_{\min}^{\delta_{\text{PA}}}(X^{\ell_2}|E^n)$ for some information leakage threshold δ_{PA}, which we establish in Theorem 29.

7.7 Analysis of PPM-MLC-Based Protocol

7.7.1 Covertness

Theorem 28. *For any $\lambda_2 > 0$, with*

$$\log h = \frac{\ell}{m_x}\chi_2(\rho_E^1\|\rho_E^0) + \sqrt{\ell}\,(2\log m_v + 3)\sqrt{\log\frac{4}{\lambda_2} + 1},$$

we have

$$\left\|\sigma_{E^nC} - \rho_{E^n}^0 \otimes \rho_C^{\text{unif}}\right\|_1 \leqslant \lambda_1 + \lambda_2 + \epsilon_{\text{IR}} + \delta_{\text{PA}}, \tag{7.16}$$

where

$$\lambda_1 = \sqrt{\frac{\ell}{2m}\chi_2(\rho_E^1\|\rho_E^0)}, \tag{7.17}$$

and ϵ_{IR} and δ_{PA} are the parameters of the information reconciliation and privacy amplification protocols, respectively, as described in Section 7.6.

Proof. Let $C = (C', C'')$ where C' contains (F, Z) and the value of hash function for privacy amplification, and C'' denotes the remaining public communication. By the triangle inequality, we have

$$\left\|\sigma_{E^nC} - \rho_{E^n}^0 \otimes \rho_C^{\text{unif}}\right\|_1 \leqslant \left\|\sigma_{E^nC} - \sigma_{E^nC'} \otimes \rho_{C''}^{\text{unif}}\right\|_1 + \left\|\sigma_{E^nC'} \otimes \rho_{C''}^{\text{unif}} - \rho_{E^n}^0 \otimes \rho_C^{\text{unif}}\right\|_1 \tag{7.18}$$

$$= \left\|\sigma_{E^nC} - \sigma_{E^nC'} \otimes \rho_{C''}^{\text{unif}}\right\|_1 + \left\|\sigma_{E^nC'} - \rho_{E^n}^0 \otimes \rho_{C'}^{\text{unif}}\right\|_1. \tag{7.19}$$

By our assumptions in (7.13)-(7.15) at the end of Section 7.6 and the leftover hash lemma [30], we have $\left\|\sigma_{E^nC} - \sigma_{E^nC'} \otimes \rho_{C''}^{\text{unif}}\right\|_1 \leqslant \epsilon_{\text{IR}} + \delta_{\text{PA}}$. We now consider the second term, i.e.,

$\left\|\sigma_{E^nC'} - \rho^0_{E^n} \otimes \rho^{\text{unif}}_{C'}\right\|_1$. Note first that, by monotonicity of the trace norm,

$$\left\|\sigma_{E^n} - \rho^0_{E^n}\right\|_1 = \left\|\mathcal{U}_{Q^n\to E^n}(\widetilde{\sigma}_{Q^nC'}) - \mathcal{U}_{Q^n\to E^n}\left(\left(\rho^0_Q\right)^{\otimes n} \otimes \rho^{\text{unif}}_{C'}\right)\right\|_1 \tag{7.20}$$

$$\leqslant \left\|\widetilde{\sigma}_{Q^nC'} - \left(\rho^0_Q\right)^{\otimes n} \otimes \rho^{\text{unif}}_{C'}\right\|_1. \tag{7.21}$$

We upper-bound the above term in two steps. Introducing an intermediate state

$$\rho^{\text{PPM}}_{Q^n} \triangleq \frac{1}{{m_x}^\ell {m_v}^\ell} \sum_{x^\ell, v^\ell} \mathcal{E}^{\otimes n}_{Q\to Q}(\rho^{x^\ell, v^\ell}_{Q^n}), \tag{7.22}$$

which is the average state at the output of the probe when v^ℓ is chosen uniformly at random from $[\![1, m_x]\!]^\ell$, we have

$$\left\|\rho^{\text{PPM}}_{Q^n} \otimes \rho^{\text{unif}}_{C'} - \left(\rho^0_Q\right)^{\otimes n} \otimes \rho^{\text{unif}}_{C'}\right\| = \left\|\rho^{\text{PPM}}_{Q^n} - \left(\rho^0_Q\right)^{\otimes n}\right\| \tag{7.23}$$

$$\overset{(a)}{\leqslant} \sqrt{\frac{1}{2}\mathbb{D}\left(\rho^{\text{PPM}}_{Q^n} \middle\| \left(\rho^0_Q\right)^{\otimes n}\right)} \tag{7.24}$$

$$\overset{(b)}{\leqslant} \sqrt{\frac{\ell}{2m}\chi_2(\rho^1_Q\|\rho^0_Q)}, \tag{7.25}$$

where (a) follows from Pinsker's inequality [85, Th. 11.9.1], and (b) follows from [91, Eq. (B144)].[2] Therefore, establishing covertness amounts to proving that the state $\sigma_{Q^nC'}$ generated by the protocol is nearly identical to $\rho^{\text{PPM}}_{Q^n} \otimes \rho^{\text{unif}}_{C'}$. This problem is known as quantum channel resolvability, and the minimum number of bits $\log h$ required is approximately equal to the Holevo information [37, Lemma 9.2]. Recall that $\mathcal{F}$ is a regular two-universal family of hash functions from $[\![1, m_v]\!]^\ell$ to $\mathcal{Z} \triangleq [\![1, {m_v}^\ell/h]\!]$.

[2] In the classical setting, the authors of [46] showed the upper-bound with a factor of $1/2$ on the right hand side. While we conjecture that an extension of such upper-bound to the quantum setting is possible, we could only prove the upper-bound without the factor $1/2$.

Let us define

$$\widetilde{\rho}^{v}_{Q^m} \triangleq \frac{1}{m_x}\sum_x |\mathrm{PPM}, d(x,v)\rangle\langle \mathrm{PPM}, d(x,v)|, \tag{7.26}$$

$$\widetilde{\rho}^{v^\ell}_{Q^n} \triangleq \widetilde{\rho}^{v_1}\otimes\cdots\otimes\widetilde{\rho}^{v_\ell}, \tag{7.27}$$

$$\widetilde{\rho}_{VQ^m} \triangleq \frac{1}{m_v}\sum_v |v\rangle\langle v|_V \otimes \widetilde{\rho}^{v}_{Q^m}, \tag{7.28}$$

$$\rho_{VQ^m} \triangleq \mathcal{E}^{\otimes m}_{Q\to Q}(\widetilde{\rho}_{VQ^m}). \tag{7.29}$$

By Lemma 53 in Appendix 7.B,

$$\left\|\sigma_{Q^nC'} - \rho^{\mathrm{PPM}}_{Q^n}\otimes\rho^{\mathrm{unif}}_{C'}\right\|_1 = \frac{1}{|\mathcal{F}|}\frac{1}{|\mathcal{Z}|}\sum_{f\in\mathcal{F},z\in\mathcal{Z}}\left\|\frac{1}{h}\sum_{v^\ell\in f^{-1}(z)}\mathcal{E}^{\otimes n}_{Q\to Q}\left(\rho^{v^\ell}_{Q^n}\right) - \rho^{\mathrm{PPM}}_{Q^n}\right\|_1 \tag{7.30}$$

$$\leqslant \lambda_2, \tag{7.31}$$

provided that

$$\log h \geqslant \log\left|\mathcal{V}^\ell\right| - \mathbb{H}^{\frac{\lambda}{4}}_{\min}(V^\ell|Q^n)_{\rho^{\otimes\ell}} + 2\log\frac{2}{\lambda_2}, \tag{7.32}$$

and $|\mathcal{V}|^\ell$ is divisible by h.[3] Applying [30, Corollary 3.3.7], we simplify the condition on $\log h$ by noting that

$$\log\left|\mathcal{V}^\ell\right| - \mathbb{H}^{\frac{\lambda}{4}}_{\min}(V^\ell|Q^n)_{\rho^{\otimes\ell}} \leqslant \log\left|\mathcal{V}^\ell\right| - \ell\left(\mathbb{H}(V|Q)_\rho - (2\mathbb{H}_{\max}(V)_\rho + 3)\sqrt{\frac{\log\frac{4}{\lambda_2}+1}{\ell}}\right) \tag{7.33}$$

$$\overset{(a)}{=} \ell\mathbb{I}(V;Q^m)_\rho + \sqrt{\ell}\,(2\log m_v + 3)\sqrt{\log\frac{4}{\lambda_2}+1}, \tag{7.34}$$

[3] For any h', we can choose h such that $|\mathcal{V}|\,h' \geqslant h \geqslant h'$ and $|\mathcal{V}|^\ell$ is divisible by h; hence, this condition adds at most $\log|\mathcal{V}| = O(\log n)$ of penalty to $\log h$.

where (a) follows since ρ_V is a mixed state. We also further upper-bound $\mathbb{I}(V;Q^m)_\rho$ by

$$\begin{aligned}
\mathbb{I}(V;Q^m)_\rho &= \mathbb{D}(\rho_{VQ^m}\|\rho_V \otimes \rho_{Q^m}) && (7.35)\\
&= \mathbb{D}\big(\rho_{VQ^m}\|\rho_V \otimes \big(\rho_Q^0\big)^{\otimes m}\big) - \mathbb{D}\big(\rho_{Q^m}\|\big(\rho_Q^0\big)^{\otimes m}\big) && (7.36)\\
&\leqslant \mathbb{D}\big(\rho_{VQ^m}\|\rho_V \otimes \big(\rho_Q^0\big)^{\otimes m}\big) && (7.37)\\
&= \frac{1}{m_v}\sum_{v\in\mathcal{V}} \mathbb{D}\big(\rho_{Q^m}^v\|\big(\rho_Q^0\big)^{\otimes m}\big) && (7.38)\\
&\overset{(a)}{=} \mathbb{D}\big(\rho_{Q^m}^1\|\big(\rho_Q^0\big)^{\otimes m}\big) && (7.39)\\
&= \mathbb{D}\big(\mathcal{E}_{Q\to Q}^{\otimes m_x}\big(\rho_{Q^{m_x}}^{\mathrm{PPM}}\big)\,\|\,\big(\rho_Q^0\big)^{\otimes m_x}\big) && (7.40)\\
&\overset{(b)}{\leqslant} \frac{1}{m_x}\chi_2(\rho_Q^1\|\rho_Q^0), && (7.41)
\end{aligned}$$

where (a) follows from the symmetry in the definition of $\rho_{Q^m}^v$, and (b) follows from [91, Eq. (B144)].

□

7.7.2 Security

The objective of this section is to lower bound the smooth min-entropy of Alice's data X^ℓ given Eve's observations. We first recall that, by our discussion in Section 7.6, we can assume that Alice prepares $\widetilde{\sigma}_{XQ^{m_x}}^{\otimes\ell}$ where

$$\widetilde{\sigma}_{XQ^{m_x}} \triangleq \frac{1}{m_x}\sum_{x=1}^{m_x} |x\rangle\langle x|_X \otimes \widetilde{\sigma}_{Q^{m_x}}^x, \tag{7.42}$$

$$\widetilde{\sigma}_{Q^{m_x}}^x \triangleq |0\rangle\langle 0|^{\otimes x-1} \otimes |\phi\rangle\langle\phi| \otimes |0\rangle\langle 0|^{\otimes m_x - x}, \tag{7.43}$$

and sends $\widetilde{\sigma}_{Q^{m_x}}^{\otimes\ell}$ over the quantum channel to Bob. We assume that Eve applies the same unitary $U_{Q^{m_x}\to Q^{m_x}E^m}$ on each PPM symbol. We do not prove the security against a coherent attack, but the techniques developed in [92, 93, 94] might be applicable to our problem. We now introduce some notation, which is summarized in Fig. 7.7.1. Let us define

Figure 7.7.1: Notations for secrecy analysis

$$\widetilde{\tau}^x_{Q^{m_x}} \triangleq \mathcal{E}^{m_x}_{Q\to Q}\left(\widetilde{\sigma}^x_{Q^{m_x}}\right), \tag{7.44}$$

$$\tau^x_{Q^{m_x}E^m} \triangleq U_{Q^{m_x}\to Q^{m_x}\to E^m}\widetilde{\tau}^x_{Q^{m_x}}U^\dagger_{Q^{m_x}\to Q^{m_x}\to E^m}, \tag{7.45}$$

$$\tau_{XQ^{m_x}E^m} \triangleq \frac{1}{m_x}\sum_{x=1}^{m_x}|x\rangle\langle x|_X \otimes \tau^x_{Q^{m_x}E^m}, \tag{7.46}$$

$$\sigma_{X^\ell Q^{m_x\ell}E^n} \triangleq \tau^{\otimes\ell}_{XBE^m}. \tag{7.47}$$

We also define

$$\widetilde{\rho}^0_{Q^{m_x}} \triangleq \left(\mathcal{E}_{Q\to Q}\left(|0\rangle\langle 0|\right)\right)^{\otimes m_x}, \tag{7.48}$$

$$\rho^0_{Q^{m_x}E^m} \triangleq U_{Q^{m_x}\to Q^{m_x}\to E^m}\widetilde{\rho}^0_{Q^{m_x}}U^\dagger_{Q^{m_x}\to Q^{m_x}\to E^m}. \tag{7.49}$$

Theorem 29. *We have*

$$\frac{1}{\ell}\mathbb{H}^\delta_{\min}(X^\ell|E^n)_\sigma \geqslant \log m_x - \mathbb{D}\big(\rho^1_Q\|\rho^0_Q\big) + \frac{1}{m_x}\sum_x \log\left(1-\eta^x\right) - \left(2\log m_x + 3\right)\sqrt{\frac{\log\frac{1}{\delta}+1}{\ell}}, \tag{7.50}$$

for all unitaries U and for all η^x such that

$$F(\tau^x_{Q^{m_x}}, \rho^0_{Q^{m_x}}) \leqslant \aleph(\lambda^x, F(|0\rangle\langle 0|, |\phi\rangle\langle\phi|)) - 2\sqrt{1-F(|0\rangle\langle 0|, |\phi\rangle\langle\phi|)}\delta - \delta^2, \tag{7.51}$$

where

$$\aleph(x,y) \triangleq 1 - \frac{2\sqrt{1-y}x + x^2}{y} - 2\sqrt{\frac{2\sqrt{1-y}x + x^2}{y}}x - x^2, \tag{7.52}$$

$$\lambda^x \triangleq 2\delta_0 + \sqrt{\eta^x + 4\sqrt{\eta^x}\delta_1 + 4(\delta_1)^2}, \tag{7.53}$$

$$\delta_1 \triangleq C(|\phi\rangle\langle\phi|, \rho_Q^1), \tag{7.54}$$

$$\delta_0 \triangleq C(|0\rangle\langle 0|, \rho_Q^0), \tag{7.55}$$

$$\delta \triangleq \delta_0 + \delta_1. \tag{7.56}$$

Remark 13. *The right hand side of* (7.50) *only depends on quantities that are either specified by the protocol and the probe, or could be calculated from Alice's and Bob's observations.*

Remark 14. *The difficulty in obtaining a bound on the adversary's information is the following. Note first that, as detailed in Chapter 6, reverse reconciliation only leads to a positive covert throughput if Eve's and Bob's observations are independent when* $|0\rangle$ *is sent. This is unfortunately not the case when the channel is a beam-splitter. To the best of our knowledge, there exist two standard methods to bound Eve's information for continuous variable QKD protocols. The first method leverages the optimality of Gaussian attack, which results in a sub-optimal bound on Eve's information for discrete-variable protocols. Since Alice's measurement is not Gaussian (in the entanglement-based version), it is not straightforward to calculate the bound for forward reconciliation protocols. The second method exploits entropic uncertainty relations, which would require finding an entanglement-based version with two different measurements at Alice. We could not find such version of our specific quantum state distribution.*

Remark 15. *Note that, in the absence of the adversary,* $\mathbb{I}(X;Q^{m_x})_\sigma = \mathbb{D}\big(\mathcal{N}(\rho_Q^1)\|\mathcal{N}(\rho_Q^0)\big) +$

$O(1/m_x)$ *[46]. Excluding finite-length effects, we achieve positive covert throughput when,*

$$\mathbb{D}\big(\rho_Q^1\|\rho_Q^0\big) - \mathbb{D}\big(\mathcal{N}(\rho_Q^1)\|\mathcal{N}(\rho_Q^0)\big) \leqslant \frac{1}{m_x}\sum_x \log(1-\eta_x). \tag{7.57}$$

This inequality holds when $\eta_x > 0$ and $\mathcal{N}$ is close to a noiseless channel.

We now state a general upper bound for the relative entropy between the outputs of the complementary channel for two fixed states.

Theorem 30. *Let A and B be two possibly infinite dimensional quantum systems such that system A is a composition of two sub-systems A' and A''. Let ρ_A^0 and ρ_A^1 be in $\mathcal{D}(\mathcal{H}_A)$ such that for two pure states $|\phi^0\rangle_{A'}$ and $|\phi^1\rangle_{A'}$ in $\mathcal{H}_{A'}$ and a mixed state $\nu_{A''}$ in $\mathcal{D}(\mathcal{H}_{A''})$, we have $C(\phi_{A'}^x \otimes \nu_{A''}, \rho_A^x) \leqslant \delta_x$. Let $\mathcal{N} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$ be a quantum channel with a complementary channel $\mathcal{E} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_E)$. Suppose that $\eta > 0$ satisfies*

$$F(\mathcal{N}(\rho_A^1), \mathcal{N}(\rho_A^0)) \leqslant \aleph(\lambda, F(\phi_{A'}^1, \phi_{A'}^0)) - 2\sqrt{1 - F(\phi_{A'}^1, \phi_{A'}^0)}\delta - \delta^2 \tag{7.58}$$

where $\lambda \triangleq 2\delta_0 + \sqrt{\eta + 4\sqrt{\eta}\delta_1 + 4\delta_1^2}$, $\delta \triangleq \delta_0 + \delta_1$.

We then have

$$\mathbb{D}\big(\mathcal{E}(\rho_A^1)\|\mathcal{E}(\rho_A^0)\big) \leqslant \mathbb{D}\big(\rho_A^1\|\rho_A^0\big) + \log(1-\eta). \tag{7.59}$$

Proof. See Appendix 7.D. □

Proof of Theorem 29. By [30, Corollary 3.3.7], we have

$$\frac{1}{\ell}\mathbb{H}_{\min}^{\epsilon}\left(X^{\ell}|E^n\right)_{\sigma} \geqslant \mathbb{H}(X|E^m)_{\tau} - \left(2\mathbb{H}_{\max}(X)_{\tau} + 3\right)\sqrt{\frac{\log\frac{1}{\epsilon}+1}{\ell}} \tag{7.60}$$

$$= \mathbb{H}(X|E^m)_{\tau} - (2\log m_x + 3)\sqrt{\frac{\log\frac{1}{\epsilon}+1}{\ell}}. \tag{7.61}$$

Furthermore,

$$\mathbb{H}(X|E^m)_\tau = \mathbb{H}(X)_\tau - \mathbb{I}(X;E^m)_\tau = \log m_x - \mathbb{I}(X;E^m)_\tau. \tag{7.62}$$

Note now that

$$\begin{aligned}
\mathbb{I}(X;E^m)_\tau &= \mathbb{D}(\tau_{XE^m}\|\tau_X \otimes \tau_{E^m}) && (7.63)\\
&= \mathbb{D}\big(\tau_{XE^m}\|\tau_X \otimes \rho^0_{E^m}\big) - \mathbb{D}\big(\tau_{E^m}\|\rho^0_{E^m}\big) && (7.64)\\
&\leqslant \mathbb{D}\big(\tau_{XE^m}\|\tau_X \otimes \rho^0_{E^m}\big) && (7.65)\\
&= \frac{1}{m_x}\sum_{x=1}^{m_x} \mathbb{D}\big(\tau^x_{E^m}\|\rho^0_{E^m}\big). && (7.66)
\end{aligned}$$

Since η_x satisfies the condition in (7.58) by (7.51), We can apply Theorem 30 to obtain

$$\mathbb{D}\big(\tau^x_{E^m}\|\rho^0_{E^m}\big) \leqslant \mathbb{D}\big(\widetilde{\tau}^x_{Q^{m_x}}\|\widetilde{\rho}^0_{Q^{m_x}}\big) + \log\left(1-\eta_x\right). \tag{7.67}$$

Combining the above inequalities, we obtain the result. □

7.7.3 Example

We present here an experimental setup over which our proposed scheme could be executed. As illustrated in Fig. 7.7.2, Alice's transmitter is a laser whose output is a single-mode bosonic system. The idle state is $|0\rangle$ and we choose a *coherent* state $|\alpha\rangle$ as the non-idle state. The probe and the honest channel are both beam-splitters with transmissivity τ_E and τ_N, respectively, and excess noise $\overline{n}_E$ and $\overline{n}_N$, respectively. In Fig. 7.7.3, we plot the number of bits per PPM symbol versus τ_N for $\tau_E = 0.9994$, $\alpha = 0.6$, $\overline{n}_E = 11$, and $\overline{n}_N = 0.01$. For these parameters, we also have $\chi_2(\mathcal{E}(|0\rangle\langle 0|)\|\mathcal{E}(|\alpha\rangle\langle\alpha|)) = 59881934$, which controls the covertness through Eq. (7.17). This example shows the possibility of covert and secret key expansion over a practical quantum channel although the efficiency is very low.

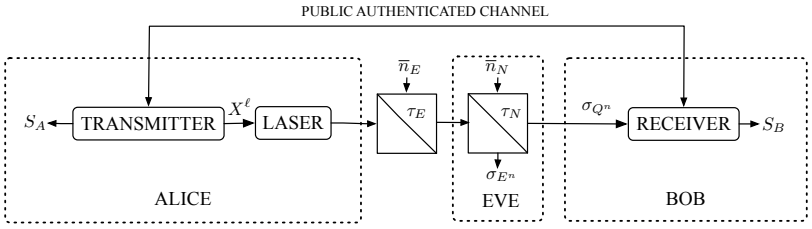

Figure 7.7.2: Experimental setup for our protocol.

Conclusion

We have developed a protocol based on PPM and MLC for the expansion of a secret key, for which we established information-theoretic secrecy of the generated keys and covertness of the protocol with respect to an adversary restricted to observing the output of a probe outside its control but otherwise only limited by the laws of quantum mechanics. We have also demonstrated the performance of our protocol for a bosonic channel. Although the range of channel parameters is narrow and the efficiency is very low, this example shows the possibility of covert QKD in settings not envisioned earlier. We believe that two factors cause the low-efficiency of our protocol: the stringent constraint of covertness and our sub-optimal bounds on Eve's information. The former factor is, in our opinion, more crucial because the optimal throughput under covertness constraint is negligible even with the knowledge Eve's attack (see Chapter 6). Using many optical mode at once [95] and developing tighter bound for Eve's information could mitigate this low-efficiency.

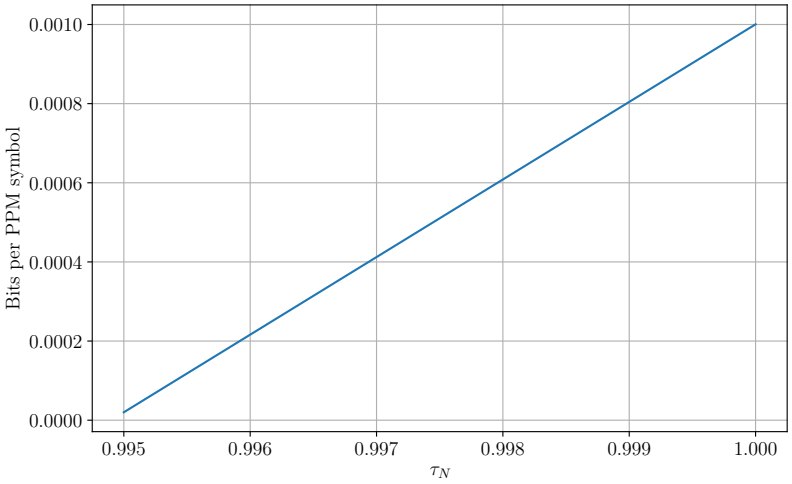

Figure 7.7.3: Achievable number of key bits per PPM symbol.

APPENDIX

7.A Proof of Theorem 27

We first prove a quantum counterpart of [70, Lemma 2.2].

Lemma 52. *Let ρ_{AB} be a bipartite state and $\mathcal{E}_{A\to AF}$ be a quantum channel. We then have*

$$\mathbb{I}(A;B)_{\rho} \geqslant \mathbb{I}(A;B|F)_{\rho'}, \tag{7.68}$$

where $\rho'_{ABF} \triangleq (\mathcal{E}_{A\to AF}\otimes \mathrm{id}_B)(\rho_{AB})$.

Proof. We have

$$\begin{aligned}
\mathbb{I}(A;B|F)_{\rho'} &= \mathbb{H}(B|F)_{\rho'} - \mathbb{H}(B|FA)_{\rho'} && (7.69)\\
&\overset{(a)}{\leqslant} \mathbb{H}(B)_{\rho'} - \mathbb{H}(B|FA)_{\rho'} && (7.70)\\
&= \mathbb{I}(AF;B)_{\rho'} && (7.71)\\
&= \mathbb{D}(\rho'_{ABF}\|\rho'_{AF}\otimes\rho') && (7.72)\\
&\overset{(b)}{\leqslant} \mathbb{D}(\rho_{AB}\|\rho_A\otimes\rho_B) = \mathbb{I}(A;B)_{\rho}, && (7.73)
\end{aligned}$$

where (a) follows from the sub-additivity of the von Neumann entropy, and (b) follows from the data processing inequality. □

Proof of Theorem 27. Let $\widetilde{\sigma}_{AQ^n}$ be the state initially prepared by Alice such that $\left\|\widetilde{\sigma}_{Q^n} - |0\rangle\langle 0|^{\otimes n}\right\|_1 \leqslant \mu$. We then have by [85, Th. 9.3.1]

$$F(\widetilde{\sigma}_{Q^n}, |0\rangle\langle 0|^{\otimes n}) \geqslant 1-\mu. \tag{7.74}$$

Let $\widetilde{\sigma}_{RAQ^n}$ be a purification of $\widetilde{\sigma}_{AQ^n}$. By Uhlmann's theorem, there exists a unit vector

$|\phi\rangle_{RA}$ such that

$$F(\widetilde{\sigma}_{RAQ^n}, \phi_{RA} \otimes |0\rangle\langle 0|^{\otimes n}) \geqslant 1 - \mu. \tag{7.75}$$

Let $\widetilde{\tau}_{AQ^n} \triangleq \operatorname{tr}_R(|\phi\rangle\langle\phi|_{RA}) \otimes |0\rangle\langle 0|^{\otimes n}$ and $\tau_{CS_AS_BE^n}$ be the output of the protocol if Alice initially prepares $\widetilde{\tau}_{AQ^n}$ instead of $\widetilde{\sigma}_{AQ^n}$. By monotonicity of the fidelity, we have $F(\widetilde{\tau}_{AQ^n}, \widetilde{\sigma}_{AQ^n}) \geqslant 1-\mu$, and therefore, $\|\widetilde{\tau}_{AQ^n} - \widetilde{\sigma}_{AQ^n}\|_1 \leqslant \sqrt{\mu}$. By the data processing inequality, we also have $\|\tau_{CS_AS_BE^n} - \sigma_{CS_AS_BE^n}\|_1 \leqslant \sqrt{\mu}$. This implies that $\mathbb{P}(S_A \neq S_B)_\tau \leqslant \epsilon + \sqrt{\mu}$. We can write the number of generated bits as

$$\begin{aligned}
K &= \mathbb{H}(S_A)_\sigma + \mathbb{D}\big(\sigma_{S_A} \| \rho_{S_A}^{\text{unif}}\big) && (7.76)\\
&\overset{(a)}{\leqslant} \mathbb{H}(S_A)_\sigma + \big\|\sigma_{S_A} - \rho_{S_A}^{\text{unif}}\big\|_1 K && (7.77)\\
&\overset{(b)}{\leqslant} \mathbb{H}(S_A)_\sigma + \big\|\sigma_{S_AC} - \rho_{S_A}^{\text{unif}} \otimes \sigma_C\big\|_1 K && (7.78)\\
&\leqslant \mathbb{H}(S_A)_\sigma + \delta K && (7.79)\\
&\overset{(c)}{\leqslant} \mathbb{H}(S_A)_\tau + \sqrt{\mu} K + \mathbb{H}_b(\sqrt{\mu}) + \delta K && (7.80)\\
&= \mathbb{H}(S_A|C)_\tau + \mathbb{I}(S_A;C)_\tau + \sqrt{\mu} K + \mathbb{H}_b(\sqrt{\mu}) + \delta K, && (7.81)
\end{aligned}$$

where (a) follows from [96, Eq. (360)], (b) follows from the data processing inequality, and (c) follows from Fannes' inequality. By [85, Exercise 11.10.2], we also have

$$\mathbb{I}(S_A;C)_\tau \leqslant \mathbb{I}(S_A;C)_\sigma + 3\sqrt{\mu} K + 2(1+\sqrt{\mu})\mathbb{H}_b\left(\frac{\sqrt{\mu}}{1+\sqrt{\mu}}\right). \tag{7.82}$$

Writing the classical state σ_{S_AC} as $\sum_c P_C(c)\sigma^c_{S_A} \otimes |c\rangle\langle c|$, we have

$$\begin{aligned}
\mathbb{I}(S_A;C)_\sigma &\leqslant \mathbb{D}\big(\sigma_{S_AC} \| \rho^{\text{unif}}_{S_A} \otimes \sigma_C\big) && (7.83)\\
&= \sum_c P_C(c) \mathbb{D}\big(\sigma^c_{S_A} \| \rho^{\text{unif}}_{S_A}\big) && (7.84)\\
&\overset{(a)}{\leqslant} \sum_c P_C(c) \big\|\sigma^c_{S_A} - \rho^{\text{unif}}_{S_A}\big\|_1 K && (7.85)\\
&= \big\|\sigma_{S_AC} - \rho^{\text{unif}}_{S_A} \otimes \sigma_C\big\|_1 K && (7.86)\\
&\leqslant \delta K, && (7.87)
\end{aligned}$$

where (a) follows from [96, Eq. (360)].

Using Fano's inequality, we obtain

$$\begin{aligned}
\mathbb{H}(S_A|C)_\tau &\leqslant \mathbb{I}(S_A;S_B|C)_\tau + \mathbb{H}_b\left(\epsilon + \sqrt{\mu}\right) + \left(\epsilon + \sqrt{\mu}\right) K && (7.88)\\
&\overset{(a)}{\leqslant} \mathbb{I}(A;Q^n)_\tau + \mathbb{H}_b\left(\epsilon + \sqrt{\mu}\right) + \left(\epsilon + \sqrt{\mu}\right) K && (7.89)\\
&\overset{(b)}{\leqslant} \mathbb{I}(A;Q^n)_{\widetilde{\tau}} + \mathbb{H}_b\left(\epsilon + \sqrt{\mu}\right) + \left(\epsilon + \sqrt{\mu}\right) K && (7.90)\\
&\overset{(c)}{=} \mathbb{H}_b\left(\epsilon + \sqrt{\mu}\right) + \left(\epsilon + \sqrt{\mu}\right) K, && (7.91)
\end{aligned}$$

where (a) follows from using Lemma 52 for each use of the public channel, (b) follows from data processing inequality, and (c) follows since $\widetilde{\tau}_{AQ^n} = \widetilde{\tau}_A \otimes \widetilde{\tau}_{Q^n}$. Combining (7.81), (7.82), (7.87), and (7.91), we obtain the desired bound. □

7.B A Quantum Resolvability Result

We prove a quantum channel resolvability result based on the privacy amplification result Lemma 7 in Chapter 1, the classical counter-part of which was proved in [21]. Note that we cannot use the standard quantum resolvability result Lemma 6 in Chapter 1 since the bound therein depends on the dimension of the output space, which itself grows exponentially for $v^\ell \mapsto \rho^{v^\ell}_{Q^n}$.

Lemma 53. *Let $\rho_{XA} = \sum_{x\in\mathcal{X}} \frac{1}{|\mathcal{X}|}|x\rangle\langle x| \otimes \rho_A^x$ be a cq state on $\mathcal{H}_X \otimes \mathcal{H}_A$. Let $\delta > 0$, h be a positive integer such that $|\mathcal{X}|$ is divisible by h and*

$$\log h \geqslant \log|\mathcal{X}| - \mathbb{H}_{\min}^{\frac{\delta}{4}}(X|A)_\rho + 2\log\frac{2}{\delta}. \tag{7.92}$$

For a regular two-universal family of hash functions $\mathcal{F}$ from $\mathcal{X}$ to $\mathcal{Z}$, we have

$$\frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}}\frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}}\left\|\rho_A - \frac{1}{h}\sum_{x\in f^{-1}(z)}\rho_A^x\right\|_1 \leqslant \delta. \tag{7.93}$$

In particular, there exists a function $g : [\![1, h]\!] \to \mathcal{X}$ such that

$$\left\|\rho_A - \frac{1}{h}\sum_{r=1}^{h}\rho_A^{g(r)}\right\|_1 \leqslant \delta. \tag{7.94}$$

Proof. Let us define $\mathcal{Z} \triangleq [\![1, \frac{|\mathcal{X}|}{h}]\!]$. By Proposition 1 and Lemma 7 in Chapter 1,

$$\frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}}\left\|(\mathcal{E}_{X\to Z}^f \otimes \mathrm{id}_A)(\rho_{XA}) - \rho_Z^{\mathrm{unif}} \otimes \rho_A\right\|_1 \leqslant \inf_{\epsilon\geqslant 0}\left[2\epsilon + 2^{-\frac{1}{2}\left(\mathbb{H}_{\min}^{\epsilon}(X|A)_\rho - \log|\mathcal{Z}|\right)}\right]. \tag{7.95}$$

By definition of $\mathcal{E}_{X\to Z}^f$ and ρ_{XA}, we have

$$(\mathcal{E}_{X\to Z}^f \otimes \mathrm{id}_A)(\rho_{XA}) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}|f(x)\rangle\langle f(x)| \otimes \rho_A^x \tag{7.96}$$

$$= \frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}}|z\rangle\langle z| \otimes \left(\frac{1}{h}\sum_{x\in f^{-1}(z)}\rho_A^x\right). \tag{7.97}$$

Therefore, we have

$$\left\|(\mathcal{E}^f_{X\to Z}\otimes \mathrm{id}_A)(\rho_{XA})-\rho_Z^{\mathrm{unif}}\otimes\rho_A\right\|_1 = \left\|\frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}}|z\rangle\langle z|\otimes\left(\frac{1}{h}\sum_{x\in f^{-1}(z)}\rho_A^x-\rho_A\right)\right\|_1 \tag{7.98}$$

$$= \frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}}\left\|\frac{1}{h}\sum_{x\in f^{-1}(z)}\rho_A^x-\rho_A\right\|_1. \tag{7.99}$$

Combining (7.95) and (7.99), we have for at least one $z\in\mathcal{Z}$ and at least one $f\in\mathcal{F}$,

$$\left\|\frac{1}{h}\sum_{x\in f^{-1}(z)}\rho_A^x-\rho_A\right\|_1 \leqslant \inf_{\epsilon\geqslant 0}\left[2\epsilon+2^{-\frac{1}{2}\left(\mathbb{H}^\epsilon_{\min}(X|A)_\rho-\log|\mathcal{Z}|\right)}\right] \tag{7.100}$$

$$\overset{(a)}{=} \inf_{\epsilon\geqslant 0}\left[2\epsilon+2^{-\frac{1}{2}\left(\mathbb{H}^\epsilon_{\min}(X|A)_\rho+\log h-\log|\mathcal{X}|\right)}\right] \overset{(a)}{\leqslant} \delta, \tag{7.101}$$

where (a) follows from (7.92). Taking a bijection $g:[\![1,h]\!]\to f^{-1}(z)$ completes the proof. □

7.C Reducing Public Communication When m_v is a Power of a Prime

In the next lemma, we show that under our symmetry conditions on $\mathcal{F}$ and ρ_{XA}, the choice of z does not matter.

Lemma 54. *Suppose that for all $f\in\mathcal{F}$, $z,z'\in\mathcal{Z}$, there exist a bijection $\phi:\mathcal{X}\to\mathcal{X}$ and unitary U acting on $\mathcal{H}_A$ (depending on z, z', and f) such that*

$$\phi(f^{-1}(z)) = f^{-1}(z') \tag{7.102}$$

$$\rho_A^{\phi(x)} = U\rho_A^x U^\dagger. \tag{7.103}$$

We then have

$$\left\| \frac{1}{h} \sum_{x \in f^{-1}(z)} \rho_A^x - \rho_A \right\|_1 = \left\| \frac{1}{h} \sum_{x \in f^{-1}(z')} \rho_A^x - \rho_A \right\|_1 . \tag{7.104}$$

Proof. Note that

$$\left\| \frac{1}{h} \sum_{x \in f^{-1}(z)} \rho_A^x - \rho_A \right\|_1 = \left\| U \left(\frac{1}{h} \sum_{x \in f^{-1}(z)} \rho_A^x - \rho_A \right) U^\dagger \right\|_1 \tag{7.105}$$

$$= \left\| \frac{1}{h} \sum_{x \in f^{-1}(z)} U \rho_A^x U^\dagger - U \rho_A U^\dagger \right\|_1 \tag{7.106}$$

$$= \left\| \frac{1}{h} \sum_{x \in f^{-1}(z)} \rho_A^{\phi(x)} - U \rho_A U^\dagger \right\|_1 \tag{7.107}$$

$$= \left\| \frac{1}{h} \sum_{x \in f^{-1}(z')} \rho_A^x - U \rho_A U^\dagger \right\|_1 . \tag{7.108}$$

Moreover, we have

$$U \rho_A U^\dagger = U \left(\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \rho_A^x \right) U^\dagger = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} U \rho_A^x U^\dagger \tag{7.109}$$

$$= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \rho_A^{\phi(x)} \tag{7.110}$$

$$= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \rho_A^x = \rho_A . \tag{7.111}$$

Therefore, we obtain (7.104). □

When m_v is a power of a prime, we provide an example of two-universal hash functions satisfying the conditions of Lemma 54. We assume in this paragraph only that $\mathcal{V} = [\![0, m_v - 1]\!]$ to be consistent with the standard notation for finite fields. Note first that $\mathcal{V}^\ell$ is a field with component-wise addition modulo m_v and a multiplication operation denoted by $\odot$. We use the short-hand 0^m for the all-zero sequence of length m and $\cdot|\cdot$

for the concatenation of two sequences. For $k \in [\![1, \ell]\!]$ and $u^\ell \in \mathcal{V}^\ell$, let $f_{u^\ell}(v^\ell)$ be the first k elements of $u^\ell \odot v^\ell$. By [97, 98], $\mathcal{F} = \left\{ f_{u^\ell} : u^\ell \in \mathcal{V}^\ell \setminus \{0^\ell\} \right\}$ is a regular two-universal class of hash functions. Moreover, for any $u^\ell \in \mathcal{V}^\ell \setminus \{0\}$, $z^k, z'^k \in \mathcal{V}^k$, we define $\phi(v^\ell) = ((z'^k - z^k)|0^{\ell-k}) \odot (u^\ell)^{-1} + v^\ell$. We show that ϕ satisfies (7.102) and (7.103). Note that

$$\begin{aligned}
\phi\left(f_{u^\ell}^{-1}(z^k)\right) &= \phi\left(\left\{v^\ell : \exists r^{\ell-k} : z^k | r^{\ell-k} = u^\ell \odot v^\ell\right\}\right) && (7.112)\\
&= \left\{v^\ell + ((z'^k - z^k)|0^{\ell-k}) \odot (u^\ell)^{-1} : \exists r^{\ell-k} : z^k | r^{\ell-k} = u^\ell \odot v^\ell\right\} && (7.113)\\
&= \left\{v^\ell : \exists r^{\ell-k} : z^k | r^{\ell-k} = u^\ell \odot (v^\ell - ((z'^k - z^k)|0^{\ell-k}) \odot (u^\ell)^{-1})\right\} && (7.114)\\
&= \left\{v^\ell : \exists r^{\ell-k} : z'^k | r^{\ell-k} = u^\ell \odot v^\ell\right\} && (7.115)\\
&= f_{u^\ell}^{-1}(z'^k) && (7.116)
\end{aligned}$$

Furthermore, let U_{CS} be the unitary operation on $\mathcal{H}_Q^{\otimes m}$ corresponding to cyclic shift of length 1, i.e., $|\phi_1\rangle \otimes \cdots \otimes |\phi_m\rangle \mapsto |\phi_m\rangle \otimes |\phi_1\rangle \otimes \cdots |\phi_{m-1}\rangle$. By definition of $d(x, v)$ and $\rho_{Q^n}^{v^\ell}$, we have

$$\rho_{Q^n}^{v^\ell + v'^\ell} = \left(U_{\mathrm{CS}}^{v'_1} \otimes \cdots \otimes U_{\mathrm{CS}}^{v'_\ell}\right) \rho_{Q^n}^{v^\ell} \left(U_{\mathrm{CS}}^{v'_1} \otimes \cdots \otimes U_{\mathrm{CS}}^{v'_\ell}\right)^\dagger, \tag{7.117}$$

where $v^\ell + v'^\ell$ is modulo m_v. We therefore conclude that (7.103) holds.

7.D Proof of Theorem 30

To prove Theorem 30, we need the following tools.

Theorem 31. ([85, Theorem 12.1.1]) *Let A and B be two quantum systems. Let ρ_A^0 and ρ_A^1 be in $\mathcal{D}(\mathcal{H}_A)$ and $\mathcal{N} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$ be a quantum channel. There exists a quantum channel $\mathcal{R} : \mathcal{D}(\mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_A)$ (depending only on $\mathcal{N}$ and ρ_A^0) such that*

$$\mathbb{D}\big(\rho_A^1 \| \rho_A^0\big) - \mathbb{D}\big(\mathcal{N}(\rho_A^1) \| \mathcal{N}(\rho_A^0)\big) \geqslant -\log F(\rho_A^1, (\mathcal{R} \circ \mathcal{N})(\rho_A^1)) \tag{7.118}$$

and

$$(\mathcal{R} \circ \mathcal{N})(\rho_A^0) = \rho_A^0. \tag{7.119}$$

Lemma 55. *Let A and B be two quantum systems such that A is a composition of two sub-systems A' and A''. Let ρ_A^0 and ρ_A^1 be in $\mathcal{D}(\mathcal{H}_A)$ such that for two pure states $|\phi^0\rangle_{A'}$ and $|\phi^1\rangle_{A'}$ in $\mathcal{H}_{A'}$ and a mixed state $\nu_{A''}$ in $\mathcal{D}(\mathcal{H}_{A''})$, we have $C(\phi_{A'}^x \otimes \nu_{A''}, \rho_A^x) \leqslant \delta_x$. Let $\mathcal{N} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_A)$ be a quantum channels such that $F(\rho_A^x, \mathcal{N}(\rho_A^x)) \geqslant 1 - \epsilon_x$. We then have*

$$F(\mathcal{E}(\rho_A^1), \mathcal{E}(\rho_A^0)) \geqslant \aleph(\lambda, F(\phi_{A'}^1, \phi_{A'}^0)) - 2\sqrt{1 - F(\phi_{A'}^1, \phi_{A'}^0)}\delta - \delta^2, \tag{7.120}$$

where $\delta \triangleq \sum_x \delta_x$, $\lambda = \sum_x \sqrt{\epsilon_x + 4\sqrt{\epsilon_x}\delta_x + 4\delta_x^2}$, $\mathcal{E}$ is a complementary channel to $\mathcal{N}$.

Proof. See Appendix 7.D.1. □

We are now ready to provide the proof of Theorem 30.

Proof. By Theorem 31, there exists a channel $\mathcal{R} : \mathcal{D}(E) \to \mathcal{D}(A)$ such that

$$\mathbb{D}\big(\rho_A^1 \| \rho_A^0\big) - \mathbb{D}\big(\mathcal{E}(\rho_A^1) \| \mathcal{E}(\rho_A^0)\big) \geqslant -\log F(\rho_A^1, (\mathcal{R} \circ \mathcal{E})(\rho_A^1)) \tag{7.121}$$

$$(\mathcal{R} \circ \mathcal{E})(\rho_A^0) = \rho_A^0. \tag{7.122}$$

Let $\mathcal{U}_{A \to BE}$ be an isometric extension of $\mathcal{N}$ compatible with $\mathcal{E}$. Let $\mathcal{W}_{E \to AF}$ be an isometric extension of $\mathcal{R}$. The isometry $(\mathbf{1}_B \otimes \mathcal{W}_{E \to AF})\mathcal{U}_{A \to BE}$ is an isometric extension of $\mathcal{R} \circ \mathcal{E}$. Hence, the mapping

$$\rho \mapsto \mathrm{tr}_A \left((\mathbf{1}_B \otimes \mathcal{W}_{E \to AF})\mathcal{U}_{A \to BE} \rho ((\mathbf{1}_B \otimes \mathcal{W}_{E \to AF})\mathcal{U}_{A \to BE})^\dagger \right) \tag{7.123}$$

is a complementary channel of $\mathcal{R} \circ \mathcal{E}$ and

$$\mathrm{tr}_{AF}\left((\mathbf{1}_B \otimes \mathcal{W}_{E\to AF})\mathcal{U}_{A\to BE}\rho((\mathbf{1}_B \otimes \mathcal{W}_{E\to AF})\mathcal{U}_{A\to BE})^\dagger\right) = \mathrm{tr}_E\left(\mathcal{U}_{A\to BE}\rho\mathcal{U}_{A\to BE}\right) = \mathcal{N}(\rho) \tag{7.124}$$

Therefore, $\mathcal{N}$ is a degraded version of the complementary channel of $\mathcal{R} \circ \mathcal{E}$. Hence, by Lemma 55, we have

$$F(\mathcal{N}(\rho_A^1), \mathcal{N}(\rho_A^0)) \geqslant \aleph(\lambda', F(\phi_{A'}^1, \phi_{A'}^0)) - 2\sqrt{1 - F(\phi_{A'}^1, \phi_{A'}^0)}\delta - \delta^2 \tag{7.125}$$

where

$$\lambda' \triangleq \sum_x \left(1 - F(\rho_A^x, \mathcal{R}(\mathcal{E}(\rho_A^x))) + 4\sqrt{1 - F(\rho_A^x, \mathcal{R}(\mathcal{E}(\rho_A^x)))}\delta_x + 4\delta_x^2\right)^{\frac{1}{2}} \tag{7.126}$$

$$= 2\delta_0 + \left(1 - F(\rho_A^1, \mathcal{R}(\mathcal{E}(\rho_A^1)) + 4\sqrt{1 - F(\rho_A^1, \mathcal{R}(\mathcal{E}(\rho_A^1))}\delta_1 + 4\delta_1^2\right)^{\frac{1}{2}}. \tag{7.127}$$

By our assumption in (7.58), we have $\aleph(\lambda, F(\phi_A^1, \phi_A^0)) \geqslant \aleph(\lambda', F(\phi_A^1, \phi_A^0))$. Since $\aleph(x, y)$ is decreasing in x for positive x, we have

$$\lambda' \geqslant \lambda, \tag{7.128}$$

which yields that $1 - \eta \geqslant 1 - F(\rho_A^1, \mathcal{R}(\mathcal{E}(\rho_A^1))$. Substituting this inequality in (7.121) completes the proof of our claim.

□

7.D.1 Proof of Lemma 55

We first prove a "triangle" inequality for fidelity measure, which follows from the triangle inequality for $C(\cdot, \cdot)$.

Lemma 56. *Let $\rho, \sigma, \rho', \sigma' \in \mathcal{D}(A)$ and let $\epsilon \triangleq C(\rho, \rho') + C(\sigma, \sigma')$. We then have*

$$F(\rho, \sigma) \geqslant F(\rho', \sigma') - 2\sqrt{1 - F(\rho', \sigma')}\epsilon - \epsilon^2. \tag{7.129}$$

Proof. By the triangle inequality for $C(\cdot, \cdot)$, we have

$$C(\rho, \sigma) \leqslant C(\rho', \sigma') + C(\rho, \rho') + C(\sigma, \sigma') = C(\rho', \sigma') + \epsilon \tag{7.130}$$

This could be written as

$$\sqrt{1 - F(\rho, \sigma)} \leqslant \sqrt{1 - F(\rho', \sigma')} + \epsilon. \tag{7.131}$$

Therefore,

$$1 - F(\rho, \sigma) \leqslant 1 - F(\rho', \sigma') + 2\epsilon\sqrt{1 - F(\rho', \sigma')} + \epsilon^2, \tag{7.132}$$

which yields the desired bound. □

We now prove a result similar to Lemma 55 when ρ_A^0 and ρ_A^1 are pure.

Lemma 57. *Let A and B be finite dimensional quantum systems such that A is a composition of two sub-systems A' and A''. Let $|\phi^0\rangle_{A'}$ and $|\phi^1\rangle_{A'}$ be pure states in $\mathcal{H}_{A'}$ and $\nu_{A''}$ be a mixed state in $\mathcal{D}(\mathcal{H}_{A''})$. Let us define $\rho_A^x \triangleq \phi_{A'}^x \otimes \nu_{A''}$. Let $V : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$ be an isometry and define $\psi_{AB}^x \triangleq V \rho_A^x V^\dagger$. Let*

$$\epsilon \triangleq \sum_x C(\psi_A^x, \rho_A^x) \tag{7.133}$$

We then have

$$F(\psi_B^1, \psi_B^0) \geqslant \aleph(\epsilon, F(\phi_{A'}^1, \phi_{A'}^0)) \tag{7.134}$$

Proof. Let $|\nu\rangle_{RA''}$ be a purification of $\nu_{A''}$ and define $|\psi^x\rangle_{RAB} \triangleq \mathbf{1}_R \otimes V(|\phi^x\rangle_{A'} \otimes |\nu\rangle_{A''R})$ (which is consistent with the definition of ψ^x_{AB}). By Uhlmann's theorem, there exist isometries U^0 and U^1 from $\mathcal{H}_R$ to $\mathcal{H}_R \otimes \mathcal{H}_B$ such that

$$C(\psi^x_A, \rho^x_A) = C(\psi^x_{ABR}, \phi^x_{A'} \otimes U^x \nu_{A''R} (U^x)^\dagger) \tag{7.135}$$

Furthermore, note that

$$\begin{aligned}
&F(\phi^1_{A'}, \phi^0_{A'}) && (7.136)\\
&= F(\phi^1_{A'} \otimes \nu_{A''R}, \phi^0_{A'} \otimes \nu_{A''R}) && (7.137)\\
&\overset{(a)}{=} F(\psi^1_{ABR}, \psi^0_{ABR}) && (7.138)\\
&\overset{(b)}{\leqslant} F(\phi^1_{A'} \otimes U^1 \nu_{A''R} (U^1)^\dagger, \phi^0_{A'} \otimes U^0 \nu_{A''R} (U^0)^\dagger) + 2\sqrt{1 - F(\psi^1_{ABR}, \psi^0_{ABR})}\epsilon + \epsilon^2 && (7.139)\\
&= F(\phi^1_{A'} \otimes U^1 \nu_{A''R} (U^1)^\dagger, \phi^0_{A'} \otimes U^0 \nu_{A''R} (U^0)^\dagger) + 2\sqrt{1 - F(\phi^1_A, \phi^0_A)}\epsilon + \epsilon^2 && (7.140)\\
&= F(\phi^1_{A'}, \phi^0_{A'}) F(U^1 \nu_{A''R} (U^1)^\dagger, U^0 \nu_{A''R} (U^0)^\dagger) + 2\sqrt{1 - F(\phi^1_A, \phi^0_A)}\epsilon + \epsilon^2, && (7.141)
\end{aligned}$$

where (a) follows since $V_{A\to AB}$ is an isometry, and (b) follows from Lemma 56 Therefore, we have

$$F(U^1 \nu_{A''R} (U^1)^\dagger, U^0 \nu_{A''R} (U^0)^\dagger) \geqslant 1 - \frac{2\sqrt{1 - F(\phi^1_A, \phi^0_A)}\epsilon + \epsilon^2}{F(\phi^1_A, \phi^0_A)} \tag{7.142}$$

Using Lemma 56 again, we obtain

$$F(\psi_B^1, \psi_B^0) \tag{7.143}$$

$$\geqslant F(U^1 \nu_{A''R}(U^1)^\dagger, U^0 \nu_{A''R}(U^0)^\dagger) - 2\sqrt{1 - F(U^1 \nu_{A''R}(U^1)^\dagger, U^0 \nu_{A''R}(U^0)^\dagger)}\epsilon - \epsilon^2 \tag{7.144}$$

$$\geqslant 1 - \frac{2\sqrt{1 - F(\phi_A^1, \phi_A^0)}\epsilon + \epsilon^2}{F(\phi_A^1, \phi_A^0)} - 2\sqrt{\frac{2\sqrt{1 - F(\phi_A^1, \phi_A^0)}\epsilon + \epsilon^2}{F(\phi_A^1, \phi_A^0)}}\epsilon - \epsilon^2 \tag{7.145}$$

$$= \aleph(\epsilon, F(\phi_A^1, \phi_A^0)). \tag{7.146}$$

□

We now prove Lemma 55. Note that for

$$\lambda \triangleq C(\phi^0, \mathcal{N}(\phi^0)) + C(\phi^1, \mathcal{N}(\phi^1)), \tag{7.147}$$

we have

$$F(\mathcal{E}(\rho_A^1), \mathcal{E}(\rho_A^0)) \overset{(a)}{\geqslant} F(\mathcal{E}(\phi_A^1), \mathcal{E}(\phi_A^0)) - 2\sqrt{1 - F(\mathcal{E}(\phi_A^1), \mathcal{E}(\phi_A^0))}\delta - \delta^2 \tag{7.148}$$

$$\geqslant F(\mathcal{E}(\phi_A^1), \mathcal{E}(\phi_A^0)) - 2\sqrt{1 - F(\phi_A^1, \phi_A^0)}\delta - \delta^2 \tag{7.149}$$

$$\overset{(b)}{\geqslant} \aleph(\lambda, F(\phi_A^1, \phi_A^0)), -2\sqrt{1 - F(\phi_A^1, \phi_A^0)}\delta - \delta^2, \tag{7.150}$$

where (a) follows from Lemma 56, and (b) follows from Lemma 57 Additionally, we have

$$F(\phi^x, \mathcal{N}(\phi^x)) \geqslant F(\rho^x, \mathcal{N}(\rho^x)) - 4\sqrt{1 - F(\rho^x, \mathcal{N}(\rho^x))}\delta_x - 4\delta_x^2 \tag{7.151}$$

$$\geqslant 1 - \epsilon_x - 4\sqrt{\epsilon_x}\delta_x - 4\delta_x^2, \tag{7.152}$$

for $x = 0, 1$. This implies that $\lambda \leqslant \sum_x \sqrt{\epsilon_x + 4\sqrt{\epsilon_x}\delta_x + 4\delta_x^2}$.

CHAPTER 8
QUANTUM STEGANOGRAPHY

8.1 Summary

We study several versions of a quantum steganography problem, in which two legitimate parties attempt to conceal a cypher in a quantum cover transmitted over a quantum channel without arising suspicion from a warden who intercepts the cover. In all our models, we assume that the warden has an inaccurate knowledge of the quantum channel and we formulate several variations of the steganography problem depending on the tasks used as the cover and the cypher task. In particular, when the cover task is classical communication, we show that the cypher task can be classical communication or entanglement sharing; when the cover task is entanglement sharing and the main channel is noiseless, we show that the cypher task can be randomness sharing; when the cover task is quantum communication and the main channel is noiseless, we show that the cypher task can be classical communication. In the latter case, our results improve earlier ones by relaxing the need for a shared key between the transmitter and the receiver and hold under milder assumptions on the cover quantum communication code. The content of this chapter is based on [99, 100].

8.2 Introduction

In steganography, two parties seek to embed information within an innocent-looking message without being detected by an unwanted party. The well-known example is that of two prisoners, Alice and Bob, who aim at developing an escape plan (*cyphertext*) through a permissible communication (*covertext*). The resulting message (*stegotext*), which is a combination of cyphertext, covertext, and possibly of a shared secret key, shall be made

available to a warden Willie and should be almost indistinguishable from the covertext. While this fictional example illustrates the main motivation behind the problem, the advent of the digital age has opened several real opportunities to conceal information, including the embedding of messages in digital images and texts as well as telecommunication networks. Applications of modern steganography are now numerous and range from copyright protection to malicious activities. The importance of such applications has led to the formalization of steganography using sound cryptographic principles and the development of both steganography methods and their countermeasures [10].

The classical information-theoretic limits of information-hiding and steganography have been studied using different measures of "hiding." The measures include average distortion between the covertext and the stegotext [101, 102] as well as relative entropy between the distributions of the covertext and stegotext [9, 103], which essentially controls the performance of the warden's optimal detector. More recently, these ideas have also been applied in the context of covert and stealth communications [7, 44]. The main insight derived from these works is the precise characterization of the number of covert bits that can be embedded in the covertext while remaining undetectable by Willie and of the number of secret key bits required by Alice and Bob to achieve this goal. The number of covert bits is sensitive to modeling assumptions, in particular to whether Willie knows the covertext or whether there is noise in the system. The authors of [20] have shown that reliable and covert transmission of $O(n)$ bits of information is possible in n uses of an AWGN channel when the warden has uncertainty about the noise power of the channel. The authors of [104, 105] have moreover considered covert communication when friendly nodes transmit artificial noise and have proved that covert transmission of positive rates is possible. Another situation in which covert communication with a positive rate was shown to be possible is the transmission from a relay node to a destination when the source is uncertain regarding the forwarding strategy of the relay node [53].

Concurrently, the quantum description of physical devices used in information process-

ing tasks has made us re-think communication and computation problems from two perspectives. First, one can use the limits imposed by quantum mechanics to devise enhanced solutions to hard problems in the classical world. For example, quantum key distribution offers unconditional security for classical communication while most classical solutions rely on assumptions regarding the computational power of the adversary. Second, one often encounters new challenging problems in a quantum setting, such as entanglement generation, which plays a role in intriguing applications such as quantum teleportation and superdense coding. Returning to the problem of steganography, one can extend the classical formulation to encompass both these aspects. That is, in addition to leveraging the quantum nature of the communication channel to perform classical steganography, one can ask for new paradigms to hide various quantum information processing tasks. Alice and Bob could, for instance, conceal a classical message within a quantum error-correcting code used to mitigate the quantum noise of a quantum computer. Because of the unique nature of quantum states and channels, quantum steganography is in principle richer than classical steganography [106], and much efforts have been devoted to characterize how much information can be embedded into various quantum channels with or without noise [107, 108, 23, 109, 110, 111], and to assess how much key is required to achieve the task.

We revisit here the model of quantum steganography put forward in [111, 109], which *assumes that the warden has inaccurate knowledge of what the channel is*. Specifically, we assume that the warden's knowledge of the channel is a degraded version of the real channel, which can be achieved by intentionally cascading another channel at the transmitter. We develop and analyze several quantum steganography protocols and obtain the following four results summarized in Table 8.2.1.

1. When the cover protocol consists of communicating classically over a quantum channel, we show that, in addition to the cover classical message, a cypher classical message can be transmitted (Theorem 32).

2. When the cover protocol consists of communicating classically over a quantum chan-

Table 8.2.1: Illustration of our results for quantum steganography

Cypher \ Cover	CC	ES	QC
CC	✓		✓
CRS		✓	
ES	✓		

✓ noisy main channel ✓ noiseless main channel

nel, we show that, in addition to the cover classical message, entangled qubits can be generated. (Theorem 33).

3. When the cover protocol consists of sharing entanglement and the channel is noiseless, we show that legitimate parties can share entanglement as well as classical randomness (Theorem 34).

4. When the cover protocol consists of quantum communication and the channel is noiseless, we show that, in addition to the cover quantum message, a cypher classical message can be transmitted (Theorem 35).

In all aforementioned results, the observed channel output state when the stego protocol is executed over the true channel resembles the observed state when the cover protocol is executed over the channel expected by the warden. Unlike earlier results [108, 111, 109], we show that no shared key is required to run the stego protocol when the channel is noiseless. This is achieved through the use of a random encoder obtained from privacy amplification and source coding with side information techniques similar to [112, 29]. Furthermore, we relax the assumption on the cover code in [111] that "*on a valid codeword in the QECC, the typical errors all have distinct error syndromes, and act as unitaries that move the state to a distinct, orthogonal subspace*," by relying on one-shot coding results. Our main results are not single-letterized because of the arbitrary structure of the cover code; however, we specialize our results to certain classes of codes and obtain a single-letterized expression for those examples.

The remainder of the paper is organized as follows. We introduce our notation in Sec-

tion 8.3. We formulate different information process protocols over a quantum channel and define our problem in Section 8.4. We state our main theorems in Section 8.5. We next calculate the rate of the cypher protocol for specific instances of cover protocols in Section 8.6. We finally prove the main theorems in Section 8.7.

8.3 Notation

Let $\emptyset_A$ be the channel that maps all states in $\mathcal{D}(\mathcal{H}_A)$ to the trivial state in a one-dimensional state.

Suppose that $\rho_{XB} = \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_B^x$ is a cq state. We recall two versions of Rényi quantum mutual information [37] for $a \neq 0$,

$$\aleph^a (X;B)_\rho \triangleq -\frac{1}{a} \log \left(\mathrm{tr} \left(\rho_{XB} (\rho_X \otimes \rho_B)^{\frac{a}{2}} \rho_{XB}^{-a} (\rho_X \otimes \rho_B)^{\frac{a}{2}} \right) \right), \tag{8.1}$$

$$\beth^a (X;B)_\rho \triangleq -\frac{1}{a} \log \left(\sum_x P_X(x) \mathrm{tr} \left((\rho_B^x)^{1-a} \rho_B^a \right) \right). \tag{8.2}$$

We also define the Rényi quantum entropy as $H^a(\rho) \triangleq -\frac{1}{a} \log \mathrm{tr} \left(\rho^{a+1}\right)$ [37]. These quantities are approximated by the Holevo information when ρ and $\mathcal{N}$ have a product structure and are useful to express the coding theorems for cq channels [37, 113, 36].

For a positive integer M, let $\mathcal{H}^{(M)}$ denote the M-dimensional space spanned by the orthonormal basis $\{|1\rangle, \cdots, |M\rangle\}$. We also define $\mathrm{id}^{(M)}(\rho) \triangleq \rho$ and $\overline{\mathrm{id}}^{(M)}(\rho) \triangleq \sum_{i=1}^M |i\rangle\langle i|\rho|i\rangle\langle i|$ for $\rho \in \mathcal{D}\left(\mathcal{H}^{(M)}\right)$. Furthermore, we define the perfectly entangled and the perfectly classically correlated states

$$|\Phi^{(M)}\rangle \triangleq \frac{1}{\sqrt{M}} \sum_{i=1}^M |i\rangle \otimes |i\rangle \in \mathcal{H}^{(M)} \otimes \mathcal{H}^{(M)} \tag{8.3}$$

$$\overline{\Phi}^{(M)} \triangleq \frac{1}{M} \sum_{i=1}^M |i\rangle\langle i| \otimes |i\rangle\langle i| \in \mathcal{D}\left(\mathcal{H}^{(M)} \otimes \mathcal{H}^{(M)}\right). \tag{8.4}$$

8.4 Problem Formulation

Suppose that Alice and Bob are connected by a quantum channel $\mathcal{N}_{A\to B} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ and use the channel n times to run a protocol, which could be a combination of four primary tasks (classical communication, quantum communication, randomness sharing, and entanglement sharing), as defined next.

- **Classical Communication:** Alice wishes to reliably transmit a classical message W uniformly distributed over $[\![1, M]\!]$. A code consists of a function $f : [\![1, M]\!] \to \mathcal{D}(\mathcal{H}_A^{\otimes n})$ for Alice to encode message w into an input state $\rho_{A^n}^w \triangleq f(w)$ and a POVM $\mathbf{\Lambda} = \{\Lambda^w\}_{w\in[\![1,M]\!]}$ for Bob to decode W. We call the code an $(M, \epsilon)^{\text{CC}}$ classical communication code, if we have $\frac{1}{M}\sum_{w=1}^{M} \text{tr}\left(\Lambda^w \mathcal{N}_{A\to B}^{\otimes n}(f(w))\right) \geqslant 1-\epsilon$. The induced output state is $\frac{1}{M}\sum_{w=1}^{M} \mathcal{N}_{A\to B}^{\otimes n}(f(w))$.

- **Quantum Communication:** Alice wants to transmit a quantum state ρ_W acting on an M-dimensional Hilbert space $\mathcal{H}_W \triangleq \mathcal{H}^{(M)}$. Alice encodes ρ_W using an encoder $\mathcal{E}_{W\to A^n}$ and transmits it over n uses of $\mathcal{N}_{A\to B}$. Bob decodes ρ_W by applying a decoder $\mathcal{D}_{B^n\to W}$ to his received state. A code $(\mathcal{E}_{W\to A^n}, \mathcal{D}_{B^n\to W})$ is an $(M, \epsilon)^{\text{QC}}$ code if

$$\min_{\rho_W\in\mathcal{D}(\mathcal{H}_W)} F(\rho_W, (\mathcal{D}_{B^n\to W} \circ \mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{E}_{W\to A^n})(\rho_W)) \geqslant 1-\epsilon. \tag{8.5}$$

  A more stringent notion of reliability is that the code recovers most of the error operators applied by the channel. Formally, we call a code $(\mathcal{E}_{W\to A^n}, \mathcal{D}_{B^n\to B})$ an $(M, \epsilon)_{\text{R}}^{\text{QC}}$ code, if there exists a decomposition $\mathcal{N}_{A\to B}^{\otimes n} = \widetilde{\mathcal{N}}_{A^n\to B^n} + \widetilde{\widetilde{\mathcal{N}}}_{A^n\to B^n}$ such that $\mathcal{D}_{B^n\to W} \circ \widetilde{\mathcal{N}}_{A^n\to B^n} \circ \mathcal{E}_{W\to A^n} = c\,\text{id}_W$ for $c \geqslant 1-\epsilon$. The induced output state is $\mathcal{N}_{A\to B}^{\otimes n}(\mathcal{E}_{W\to A^n}(\rho_W))$ when the message is ρ_W.

- **Randomness Sharing:** Alice and Bob desire to share a classical random variable $\overline{\Phi}^{(M)}$. Let $\mathcal{H}_{\widetilde{A}} = \mathcal{H}_{\widetilde{B}} \triangleq \mathcal{H}^{(M)}$. Alice prepares a state $\rho_{\widetilde{A}A^n}$ over the Hilbert space

$\mathcal{H}_{\widetilde{A}} \otimes \mathcal{H}_A^{\otimes n}$ and transmits ρ_{A^n} to Bob over n uses of the channel $\mathcal{N}_{A\to B}$. Bob applies a decoder $\mathcal{D}_{B^n\to\widetilde{B}}$ to his received state $\mathcal{N}_{A\to B}^{\otimes n}(\rho_{A^n})$ to obtain the state $\rho_{\widetilde{B}}$ acting on the Hilbert space $\mathcal{H}_{\widetilde{B}}$. The joint state $\rho_{\widetilde{A}\widetilde{B}} \triangleq (\mathrm{id}_{\widetilde{A}} \otimes (\mathcal{D}_{B^n\to\widetilde{B}} \circ \mathcal{N}_{A\to B}^{\otimes n}))(\rho_{\widetilde{A}A^n})$ is their final shared randomness. A code $(\rho_{\widetilde{A}A^n}, \mathcal{D}_{B^n\to\widetilde{B}})$ is called an $(M,\epsilon)^{\mathrm{RS}}$ randomness sharing code if $F(\overline{\Phi}^{(M)}, \rho_{\widetilde{A}\widetilde{B}}) \geqslant 1-\epsilon$. The induced output state is $\mathcal{N}_{A\to B}^{\otimes n}(\rho_{A^n})$.

- **Entanglement Sharing:** Alice and Bob want to share the entangled state $\Phi^{(M)}$. An $(M,\epsilon)^{\mathrm{ES}}$ code is defined in the same way as a randomness sharing protocol except that the final desired state is $\Phi^{(M)}$. The induced output state is defined similarly to that of randomness sharing.

In the resource framework formulated in [114], these four protocols correspond to the simulation of $\overline{\mathrm{id}}^{(M)}$, $\mathrm{id}^{(M)}$, $\overline{\Phi}^{(M)}$, and $\Phi^{(M)}$ with n uses of $\mathcal{N}_{A\to B}$. Alice and Bob can in principle desire to perform any combination of these four protocols over n uses of the channel $\mathcal{N}_{A\to B}$. We formalize only the combinations for which we develop results, i.e., classical communication/quantum communication, entanglement sharing/randomness sharing, and entanglement sharing/classical communication.

- **Quantum and Classical Communication:** Alice wants to transmit a quantum state ρ_W over an M-dimensional space $\mathcal{H}_W \triangleq \mathcal{H}^{(M)}$ and an independent classical message $\overline{W}$ uniformly distributed over $[\![1, \overline{M}]\!]$. When $\overline{W} = \overline{w}$, she encodes ρ_W using the encoder $\mathcal{E}_{W\to A^n}^{\overline{w}}$. Bob decodes the messages using a decoder $\mathcal{D}_{B^n\to W\overline{W}}$. The code is called an $(M, \overline{M}, \epsilon)^{\text{QC-CC}}$ code if for any ρ_W, we have

$$\frac{1}{\overline{M}} \sum_{\overline{w}} \mathrm{tr}\left(|\overline{w}\rangle\langle\overline{w}| (\mathcal{D}_{B^n\to\overline{W}} \circ \mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{E}_{W\to A^n}^{\overline{w}})(\rho_W)\right) \geqslant 1-\epsilon, \tag{8.6}$$

  and for all $\overline{w} \in [\![1, \overline{M}]\!]$, $(\mathcal{E}_{W\to A^n}^{\overline{w}}, \mathcal{D}_{B^n\to W})$ is an $(M,\epsilon)_R^{\mathrm{QC}}$ code. The induced output state is $\frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \mathcal{N}_{A\to B}^{\otimes n}(\mathcal{E}_{W\to A^n}^{\overline{w}}(\rho_W))$ when the quantum message is ρ_W.

- **Entanglement and Randomness Sharing:** Alice and Bob want to share the state

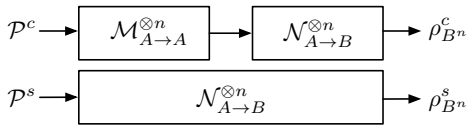

Figure 8.4.1: Willie's expectation (top) and true communication (bottom)

$\Phi^{(M)} \otimes \overline{\Phi}^{(\overline{M})}$. An $(M, \overline{M}, \epsilon)^{\text{ES-RS}}$ code is defined in the same way as a randomness sharing protocol except that the final desired state is $\Phi^{(M)} \otimes \overline{\Phi}^{(\overline{M})}$. The induced output state is defined similarly to that of randomness sharing.

- **Classical Communication and Entanglement Sharing**: Alice wants to transmit a classical message W uniformly distributed over $[\![1, M]\!]$ and share the entangled state $\Phi^{(\overline{M})}$ with Bob. Let $\mathcal{H}_{\widetilde{A}} = \mathcal{H}_{\widetilde{B}} \triangleq \mathcal{H}^{(\overline{M})}$ and $\mathcal{H}_W \triangleq \mathcal{H}^{(M)}$. A code consists of an encoder $f : [\![1, M]\!] \to \mathcal{H}_{\widetilde{A}} \otimes \mathcal{H}_{A^n}$ and a decoder $\mathcal{D}_{B^n \to W\widetilde{B}}$. Given the classical message $W = w$, Alice prepares $f(w)$ and sends the subsystem A^n over n uses of $\mathcal{N}_{A\to B}$. Bob applies $\mathcal{D}_{B^n \to W\widetilde{B}}$ to his received state. We call $(f, \mathcal{D}_{B^n \to W\widetilde{B}})$ an $(M, \overline{M}, \epsilon)^{\text{CC}-\text{ES}}$ code if

$$\frac{1}{M} \sum_w \langle w | \mathcal{D}_{B^n \to W} \circ \mathcal{N}_{A\to B}^{\otimes n}(\text{tr}_{\widetilde{A}}(f(w))) | w \rangle \geqslant 1 - \epsilon, \tag{8.7}$$

$$\left\| \frac{1}{M} \sum_w (\text{id}_{\widetilde{A}} \otimes (\mathcal{D}_{B^n \to \widetilde{B}} \circ \mathcal{N}_{A\to B}^{\otimes n}))(f(w)) - \Phi^{(\overline{M})} \right\|_1 \leqslant \epsilon. \tag{8.8}$$

The induced output state is $\frac{1}{M} \sum_{w=1}^{M} \mathcal{N}_{A\to B}^{\otimes n}(\text{tr}_{\widetilde{A}}(f(w)))$.

All these protocols can be enhanced with a shared secret classical key S uniformly distributed over $[\![1, K]\!]$, which can help Alice and Bob induce a specific output state. As depicted in Fig. 8.4.1, Willie expects Alice and Bob to execute a protocol $\mathcal{P}^c$, which is called *the cover protocol* and is known to Willie. However, Willie has an inaccurate estimation of the channel and believes that the channel between Alice and Bob is $\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A}$, which is a degraded version of the true channel $\mathcal{N}_{A\to B}$. We assume that running the protocol $\mathcal{P}^c$ induces the quantum state $\rho_{B^n}^c$ at the output of $\mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{M}_{A\to A}^{\otimes n}$. The objective is for Alice

and Bob to run a *stego protocol* $\mathcal{P}^s$, which performs the task of $\mathcal{P}^c$ together with another task and induces a state $\rho^s_{B^n}$ at the output of $\mathcal{N}^{\otimes n}_{A\to B}$ such that $\|\rho^c_{B^n} - \rho^s_{B^n}\|_1$ is small. The added tasks can be any of the tasks listed earlier. We focus on four of these as summarized in Table 8.2.1 and detailed next.

8.5 Main Results

We state our main results in this section, and all proofs are relegated to Section 8.7. We first show that if the cover protocol is a classical communication code, the stego protocol could be a classical communication code with a higher rate, equivalent to sending a cypher classical message in addition to the cover classical message.

Theorem 32 (classical communication / classical communication)**.** *Let the cover protocol be an* $(M,\epsilon)^{\text{CC}}$ *code* $(f,\mathbf{\Lambda})$ *for* $\mathcal{N}^{\otimes n}_{A\to B}\circ\mathcal{M}^{\otimes n}_{A\to A}$ *inducing the output state* $\rho^c_{B^n}$*. We define* $\rho^w_{B^n} \triangleq \mathcal{N}^{\otimes n}_{A\to B}\circ\mathcal{M}^{\otimes n}_{A\to A}(f(w))$ *for* $w\in[\![1,M]\!]$.

- *Suppose that* $\mathcal{H}_A = \mathcal{H}_B$ *and* $\mathcal{N}_{A\to B} = \text{id}_A$*, i.e., the true channel from Alice to Bob is noiseless. For any* $\zeta > 0$*, there exists an* $(M\overline{M}, \zeta + 2\sqrt{\zeta+\epsilon})^{\text{CC}}$ *stego protocol inducing the output state* $\rho^s_{B^n}$ *such that* $\|\rho^s_{B^n} - \rho^c_{B^n}\|_1 \leqslant \zeta$ *provided that*

$$\log \overline{M} \leqslant \min_{w\in[\![1,M]\!]} \sup_{a\in]0,1[} H^a(\rho^w_{B^n}) - \frac{4}{a}\log\frac{2}{\zeta}. \tag{8.9}$$

- *Suppose that the channel* $\mathcal{N}_{A\to B}$ *is noisy. Let* $\sigma^1_{XA^n},\cdots,\sigma^M_{XA^n}$ *be cq states such that upon defining* $\sigma^w_{XB^n} \triangleq (\text{id}_X\otimes\mathcal{N}^{\otimes n}_{A\to B})(\sigma^w_{XA^n})$*, we have* $\text{tr}_X(\sigma^w_{XB}) = \rho^w_{B^n}$ *for all* $w\in[\![1,M]\!]$*. Let* $\zeta,\xi\in]0,1[$ *be fixed and* $\overline{M}$ *and* $\overline{K}$ *be positive integers such that* $\log\overline{M}\leqslant\min_w\log\overline{M}_w$ *where*

$$\log\overline{M}_w \triangleq \sup_{a\in]0,1[}\left[\aleph^a\left(X;B^n\right)_{\sigma^w} - \frac{1}{a}\log\nu(\sigma^w_X\otimes\sigma^w_{B^n}) - \frac{4}{a}\log\frac{12}{\zeta}\right] \tag{8.10}$$

and

$$\log \overline{K} \geqslant \max_{w} \left\{ \inf_{a<0} [\beth^{a}\left(X;B^{n}\right)_{\sigma^{w}} + \log \nu(\rho^{w}_{B^{n}}) + \left(2 - \frac{2}{a}\right) \log \frac{12}{\xi} \right] - \log \overline{M}_{w} + 1 \right\}. \quad (8.11)$$

There exists an $(M\overline{M}, \zeta + 2\sqrt{\xi+\epsilon})^{\mathrm{CC}}$ *code with* $\log \overline{K}$ *bits of required common randomness inducing the output state* $\rho^{s}_{B^{n}}$ *such that* $\left\| \rho^{s}_{B^{n}} - \rho^{c}_{B^{n}} \right\|_{1} \leqslant \xi$.

Remark 16. *We assume for simplicity that the cover message is uniformly distributed, but the proof holds for all distributions on the cover message.*

Remark 17. *The arbitrary choice of* $\sigma^{1}_{XA^{n}}, \cdots, \sigma^{M}_{XA^{n}}$ *in the second part of Theorem 32 is an essential part of most of the channel coding results, for example the choice of the channel input state in the definition of the Holevo information of a quantum channel [85, Definition 13.3.1]. We need however an additional requirement* $\mathrm{tr}_{X}(\sigma^{w}_{XB}) = \rho^{w}_{B^{n}}$ *to control the channel output statistics.*

We next show that if the cover protocol is a classical communication code, we can use a stego protocol to share entanglement and communicate classically. We introduce the following two definitions to express our results. In the first definition we introduce a shorthand for the result of Theorem 32. It shall help us compactly state the next theorem as we use the stego protocol of Theorem 32 as a sub protocol in our stego protocol of Theorem 33.

Definition 11. *Let us fix* $\xi = \zeta$ *in the second part of Theorem 32. For an encoder* f *and positive number* ζ*, let* $\log \overline{M}^{\mathrm{CC}}(f,\zeta)$ *and* $\log K^{\mathrm{CC}}(f,\zeta)$ *be the number of bits of the cypher message and the number of required key bits, respectively, in the stego protocol of Theorem 32. Note that these quantities are well-defined, because the right hand side of* (8.9)*,* (8.10)*, and* (8.11) *only depends on* f*,* ζ*, and* ξ *when the channel is fixed.*

We next introduce a notation for the maximum amount of entanglement that can be distilled from an arbitrary shared quantum state using local operations and classical communication, known as the entanglement distillation problem.

Definition 12. *Let Alice and Bob share ρ_{AB} and $\mathcal{H}_{\widetilde{A}} = \mathcal{H}_{\widetilde{B}} \triangleq \mathcal{H}^{(M)}$. An entanglement distillation protocol consists of an encoder $\mathcal{E}_{A\to C\widetilde{A}}$ and a decoder $\mathcal{D}_{BC\to\widetilde{B}}$ such that the output of $\mathcal{E}_{A\to C\widetilde{A}}$ is always a cq state. Alice applies $\mathcal{E}_{A\to C\widetilde{A}}$ to ρ_A to obtain a cq state $\rho_{C\widetilde{A}}$ and transmits C to Bob over a noiseless channel. Bob applies $\mathcal{D}_{CB\to\widetilde{B}}$ to his subsystem B and the received classical message C. The code $(\mathcal{E}_{A\to C\widetilde{A}}, \mathcal{D}_{CB\to\widetilde{B}})$ is called an $(M, L, \rho_{AB}, \epsilon)^{\mathrm{ED}}$ code if $\dim \mathcal{H}_C = L$ and*

$$\left\| (\mathrm{id}_{\widetilde{A}} \otimes \mathcal{D}_{CB\to\widetilde{B}}) \circ (\mathcal{E}_{A\to C\widetilde{A}} \otimes \mathrm{id}_B)(\rho_{AB}) - \Phi^{(M)} \right\|_1 \leqslant \epsilon. \tag{8.12}$$

We further define $E_d(\rho_{AB}, L, \epsilon) \triangleq \max\left\{M : \exists\, (M, L, \rho_{AB}, \epsilon)^{\mathrm{ED}}\mathrm{code}\right\}$.

When ρ_{AB} is pure, it is known [115] that $\lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{\log E_d(\rho_{AB}^{\otimes n}, 2^{\Theta(\log n)}, \epsilon)}{n} = \mathbb{H}(A)_\rho$.

Theorem 33 (classical communication / entanglement sharing)**.** *Let the cover protocol be an $(M,\epsilon)^{\mathrm{CC}}$ code $(f, \mathbf{\Lambda})$ for $\mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{M}_{A\to A}^{\otimes n}$ inducing the output state $\rho_{B^n}^c$. Further assume that $f(w) = f_1(w) \otimes f_2(w)$ for two functions $f_1 : [\![1, M]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n_1}\right)$ and $f_2 : [\![1, M]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n_2}\right)$ where $n = n_1 + n_2$. Let $|\phi^w\rangle_{RA^{n_1}}$ be a purification of $\mathcal{M}_{A\to A}^{\otimes n_1}(f_1(w))$ and $\sigma_{RB^{n_1}}^w \triangleq \mathrm{id}_R \otimes \mathcal{N}_{A\to B}^{\otimes n_1}(|\phi^w\rangle\langle\phi^w|_{RA^{n_1}})$. For any $\zeta > 0$, there exists an $(M, \overline{M}, 2\zeta + 2\sqrt{\epsilon+\zeta} + 2\sqrt{\zeta + 2\sqrt{\epsilon+\zeta}})^{\mathrm{CC-ES}}$ stego protocol inducing the output state $\rho_{B^n}^s$ such that $\|\rho_{B^n}^s - \rho_{B^n}^c\|_1 \leqslant \zeta$ provided that $\overline{M} \leqslant \min_{w\in[\![1,M]\!]} E_s(|\phi\rangle_{RB^{n_1}}^w, \overline{M}^{\mathrm{CC}}(f_2,\zeta),\zeta)$.*

The stego protocol requires $\log \overline{M}^{\mathrm{CC}}(f_2,\zeta) + \log K^{\mathrm{CC}}(f_2,\zeta)$ bits of shared key.

Remark 18. *Our assumption that $f(w)$ decomposes as $f_1(w) \otimes f_2(w)$ for all w holds for common codes for classical communication over quantum channels such as [116].*

We next show that if the cover protocol is an entanglement sharing code, there exists a stego protocol that shares both entanglement and classical randomness.

Theorem 34 (entanglement sharing / classical randomness sharing)**.** *Let the cover protocol be an* $(M,\epsilon)^{\text{ES}}$ *code* $(\rho_{\widetilde{A}A^n},\mathcal{D}_{B^n\to\widetilde{B}})$ *for* $\mathcal{N}_{A\to B}^{\otimes n}\circ\mathcal{M}_{A\to A}^{\otimes n}$ *inducing the output state* $\rho_{B^n}^c$*. If* $\mathcal{H}_A=\mathcal{H}_B$ *and* $\mathcal{N}_{A\to B}=\mathrm{id}_A$*, for any* $\zeta\geqslant 0$ *and* $\overline{M}$*, there exists an* $(M,\overline{M},(\sqrt{\epsilon}+\zeta)^2)^{\text{ES-RS}}$ *stego protocol inducing the output state* $\rho_{B^n}^c$ *such that* $\rho_{B^n}^s=\rho_{B^n}^s$ *if*

$$\log\overline{M}\leqslant\sup_{a\in]0,1[}\left(H^a((\mathrm{id}_{\widetilde{A}}\otimes\mathcal{M}_{A\to A}^{\otimes n})(\rho_{\widetilde{A}A^n}))-\frac{4}{a}\log\frac{2}{\zeta}\right). \tag{8.13}$$

Finally we show that a cover protocol for quantum communication can be converted into a quantum and classical communication stego protocol.

Theorem 35 (quantum communication / classical communication)**.** *Let the cover protocol be an* $(M,\epsilon)_{\text{R}}^{\text{QC}}$ *code* $(\mathcal{E}_{W\to A^n},\mathcal{D}_{B^n\to W})$ *inducing the output state* $\rho_{B^n}^c$*. Suppose that* $\mathcal{E}_{W\to A^n}=V_{W\to A^n}\rho_W V_{W\to A^n}^\dagger$ *where* $V_{W\to A^n}$ *is an isometry. If* $\mathcal{H}_A=\mathcal{H}_B$ *and* $\mathcal{N}_{A\to B}=\mathrm{id}_A$*, for all* $\zeta>4\sqrt{\epsilon}$*, there exists an* $(M,\overline{M},\max(\zeta,\epsilon))_{\text{R}}^{\text{QC-CC}}$ *stego protocol inducing the output state* $\rho_{B^n}^s$ *such that* $\|\rho_{B^n}^s-\rho_{B^n}^c\|_1\leqslant 2\epsilon+\zeta$*,*[1] *provided that*

$$\log\overline{M}\leqslant\sup_{a\in]0,1[}H^a\left(\mathcal{M}_{A\to A}^{c\,\otimes n}\left(\mathcal{E}_{W\to A^n}\left(\frac{1}{\overline{M}}\mathbf{1}_W\right)\right)\right)-\frac{4}{a}\log\frac{2}{\zeta/2-2\sqrt{\epsilon}}, \tag{8.14}$$

where $\mathcal{M}_{A\to A}^c$ *is the complementary channel of* $\mathcal{M}_{A\to A}$*.*

8.6 Examples

8.6.1 Classical Codes with Product Structure

Definition 13. *Let* k *and* ℓ *be positive integers, and* $\rho_{A^k}^1,\cdots,\rho_{A^k}^\ell\in\mathcal{D}(\mathcal{H}_A^{\otimes k})$*. We say that an encoder* $f:[\![1,M]\!]\to\mathcal{D}(\mathcal{H}_A^{\otimes n})$ *has* a product structure with respect to $\mathcal{P}_{k,\ell}\triangleq\{\rho_{A^k}^1,\cdots,\rho_{A^k}^\ell\}$*, if* n *is divisible by* k *and for all* $w\in[\![1,M]\!]$*, we have* $f(w)=\otimes_{i=1}^{n/k}\sigma^i$ *where* $\sigma^1,\cdots,\sigma^{n/k}\in\mathcal{P}_{k,\ell}$*.*

Remark 19. *Definition 13 is useful when* $n/k\gg 1$*. Several explicit constructions of*

[1] Note that $\rho_{B^n}^s$ and $\rho_{B^n}^c$ depend on ρ_W, and this inequality should hold for all choices of ρ_W.

classical codes for quantum channels are in this regime [116]. Moreover, from the standard random coding arguments, codes with large n/k achieve the classical capacity of any quantum channel.

Considering the cover classical communication code described in Theorem 32, we simplify the expressions for the rate of the cypher message provided that the cover code has a product structure and n/k is large enough. Let $\delta > 0$ and let the classical communication code have *a product structure with respect to* $\mathcal{P}_{k,\ell}$. There exist an integer m depending on $\mathcal{P}_{k,\ell}$, $\zeta, \delta > 0$ such that if $n/k \geqslant m$ the following two propositions hold.

Proposition 3. *For a noiseless channel, the number of bits of the cypher message is at least* $\frac{n}{k}\left(\min_{\rho\in\mathcal{P}_{k,\ell}} H(\mathcal{M}_{A\to A}^{\otimes k}(\rho)) - \delta\right)$. *For a noisy channel, the number of bits of the cypher message is at least* $\frac{n}{k}\left(\inf_{\rho\in\mathcal{P}_{k,\ell}} \sup_{\sigma_{XB^k}:\mathrm{tr}_X(\sigma_{XB^k})=\rho} \mathbb{I}\big(X;B^k\big)_\sigma - \delta\right)$*, using a shared secret key of* δn *bits.*

Proposition 4. *For a noiseless channel, the number of entangled qubits that the stego protocol of Theorem 33 would generate is at least* $\frac{n}{k}\left(\min_{\rho\in\mathcal{P}_{k,\ell}} H(\mathcal{M}_{A\to A}^{\otimes k}(\rho)) - \delta\right)$. *The required number of shared secret key bits is* $O(\log n)$.

8.6.2 Gaussian States

Although we have assumed so far that all Hilbert spaces are finite dimensional, the proof of the first part of Theorem 32 carries over to infinite dimensional spaces since the left-over hash lemma still holds for such a setting. Gaussian channels form an important class of infinite dimensional channels, which models optical channels. Let A and B be single mode bosonic systems, $\mathcal{N}_{A\to B}$ be noiseless, $\mathcal{M}_{A\to A}$ be a Gaussian channel, and $f(w)$ be a Gaussian state for all w. Denoting the symplectic spectra of ρ_{B^n} by $(\nu_1^w, \cdots, \nu_n^w)$, we have [117, Eq. (108)] $H^a(\rho_{B^n}^w) = -\frac{\sum_{i=1}^n \log\left(\eta_{1+a}(\nu_i^w)\right)}{a}$, where $\eta_\alpha(x) \triangleq 2^\alpha/((x+1)^\alpha - (x-1)^\alpha)$.

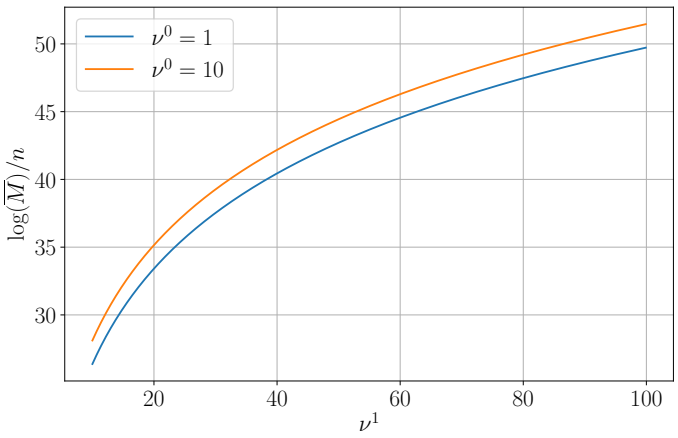

Figure 8.6.1: Rate of the cypher message vs the symplectic eigenvalues of $\mathcal{M}_{A\to A}(\rho_A^0)$ and $\mathcal{M}_{A\to A}(\rho_A^1)$, ν^0 and ν^1

The number of bits of the cypher message would then be

$$\log \overline{M} = \min_{w\in[\![1,M]\!]} \sup_{a\in]0,1[} -\frac{\sum_{i=1}^n \log\left(\eta_{1+a}(\nu_i^w)\right)}{a} - \frac{4}{a}\log\frac{2}{\zeta}. \tag{8.15}$$

We now suppose that the cover code uses a binary modulation, i.e., for two states ρ_A^0 and ρ_A^1, we have $f(w) = \otimes_{i=1}^n \rho_A^{x_{w,i}}$ for all $w \in [\![1, M]\!]$. Let ν^0 and ν^1 be the symplectic eigenvalue of $\mathcal{M}_{A\to A}(\rho_A^0)$ and $\mathcal{M}_{A\to A}(\rho_A^1)$, respectively, with $\nu^0 \leqslant \nu^1$. Upon defining $r \triangleq \min_{w\in[\![1,M]\!]} \sum_{i=1}^n x_{w,i}$, we have

$$\log \overline{M} = \sup_{a\in]0,1[} -\frac{(n-r)\log\left(\eta_{1+a}(\nu^0)\right) + r\log\left(\eta_{1+a}(\nu^1)\right)}{a} - \frac{4}{a}\log\frac{2}{\zeta}. \tag{8.16}$$

We plot the rate of the cypher message for $n = 10^6$, $r = n/2$, $\zeta = 10^{-3}$ in Fig. 8.6.1.

8.6.3 Quantum Codes of [111]

Consider a Kraus representation $\{F_j\}_{j\in\mathcal{J}}$ of $\mathcal{M}_{A\to A}$ such that $\mathrm{tr}\left(F_j^\dagger F_{j'}\right) = \mathbb{1}\{j = j'\}\, d_j$. This defines a Kraus representation $\{\mathbf{F_j}\}_{\mathbf{j}\in\mathcal{J}^n}$ for $\mathcal{M}_{A\to A}^{\otimes n}$, where $\mathbf{F_j} \triangleq F_{j_1}\otimes\cdots\otimes F_{j_n}$. Let $\mathcal{T}_\epsilon$ be the typical subset of $\mathbf{F_j} \triangleq F_{j_1}\otimes\cdots\otimes F_{j_n}$ as defined in [118]. If Π is the projector onto the sub-space of inputs defined by the code, we assume that for all $\mathbf{F_j} \in \mathcal{T}_\epsilon$, we have

$\mathbf{F_j}\Pi = p_{\mathbf{j}}\mathbf{U_j}\Pi$, where $p_{\mathbf{j}} = p_{j_1} \times \cdots \times p_{j_n}$ for a probability distribution $\{p_j\}$ on $\mathcal{J}$, and $\mathbf{U_j} = U_{j_1} \otimes \cdots \otimes U_{j_n}$ for unitaries $\{U_j\}$ on $\mathcal{H}_A$.

Proposition 5. *For all $\delta > 0$ and n large enough, we have*

$$\sup_{a\in]0,1[} H^a\left(\mathcal{M}^c_{A\to A}{}^{\otimes n}\left(\mathcal{E}_{W\to A^n}\left(\frac{1}{M}\mathbf{1}_W\right)\right)\right) \geqslant n\left(-\sum_j p_j \log p_j - \delta\right). \tag{8.17}$$

8.6.4 Random Quantum Codes

Proposition 6. *Let $\mathcal{S}$ be a random M-dimensional subspace of $\mathcal{H}_A^{\otimes n}$ distributed according to the Haar measure, and Π denote the projector onto $\mathcal{S}$. Let $\mathcal{M}_{A\to A}$ be a quantum channel with an isometric extension $V_{A\to AE}$. For all $\delta > 0$, there exists n large enough such that*

$$\mathbb{E}_{\mathcal{S}}\left(\sup_{a\in]0,1[} H^a\left(\mathcal{M}^c_{A\to A}{}^{\otimes n}\left(\frac{\Pi}{M}\right)\right)\right) \geqslant n\left(\log\min(\text{rank}\left(\text{tr}_A(VV^\dagger)\right), \text{rank}\left(\text{tr}_E(VV^\dagger)\right)\right) - \delta\right) \tag{8.18}$$

Proof. Let $|g^1\rangle, \cdots, |g^M\rangle$ be M random independent Gaussian vectors in $\mathcal{H}_A^{\otimes n}$ as defined in [119]. Since the distribution of $\text{span}(|g^1\rangle, \cdots, |g^M\rangle)$ is the same as the distribution of $\mathcal{S}$ by [119], we take $\mathcal{S} = \text{span}(|g^1\rangle, \cdots, |g^M\rangle)$. Defining $G \triangleq \sum_{j=1}^M |g^j\rangle\langle g^j|$, the vectors $\{|\phi^j\rangle_{A^n} \triangleq G^{-\frac{1}{2}}|g^j\rangle\}_{j\in[\![1,M]\!]}$ form an orthonormal basis for $\mathcal{S}$. One can check that $|\psi^j\rangle_{A^nE^n} \triangleq V_{A\to AE}|\phi^j\rangle$ has a uniform distribution over all unit vectors in $(\text{range}VV^\dagger)^{\otimes n}$. Therefore, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{S}}\left(H^a\left(\mathcal{M}^c_{A\to A}{}^{\otimes n}\left(\frac{\Pi}{M}\right)\right)\right) &\overset{(a)}{=} \mathbb{E}_{\mathcal{S}}\left(H^a\left(\frac{1}{M}\sum_{j=1}^M \mathcal{M}^c_{A\to A}{}^{\otimes n}\left(|\phi^j\rangle\langle\phi^j|\right)\right)\right) && (8.19)\\
&\geqslant \frac{1}{M}\sum_{j=1}^M \mathbb{E}_{\mathcal{S}}\left(H^a\left(\mathcal{M}^c_{A\to A}{}^{\otimes n}\left(|\phi^j\rangle\langle\phi^j|\right)\right)\right) && (8.20)\\
&= \frac{1}{M}\sum_{j=1}^M \mathbb{E}_{\mathcal{S}}\left(H^a\left(\psi^j_{E^n}\right)\right) && (8.21)\\
&\overset{(b)}{\geqslant} n\log\text{rank}\left(\text{tr}_A(VV^\dagger)\right) + n\log\text{rank}\left(\text{tr}_E(VV^\dagger)\right)
\end{aligned}$$

$$- 2 \log \left(\sqrt{\text{rank}\,(\text{tr}_A(VV^\dagger))^n} + \sqrt{\text{rank}\,(\text{tr}_E(VV^\dagger))^n} \right) \tag{8.22}$$

$$\geqslant n \log \min(\text{rank}\left(\text{tr}_A(VV^\dagger)\right), \text{rank}\left(\text{tr}_E(VV^\dagger)\right)) - \log 2, \tag{8.23}$$

where (a) follows from the concavity of Rényi entropy, and (b) follows from the bound in the proof of [120, Lemma III.1]. □

8.7 Proofs

8.7.1 One-Shot Results

In this section, we develop one-shot coding results stated in terms of Rényi mutual information. We shall specialize them to prove our main results in Section 8.7.2. We first derive an achievability result stating that there exists a classical communication code for a cq channel inducing a pre-specified state at the output. Our proof is based on combining quantum channel coding and channel resolvability results.

Lemma 58. *Let $\rho_{XB} = \sum_{x \in \mathcal{X}} P_X(x)|x\rangle\langle x| \otimes \rho_B^x$ be a cq state. Let M and K be positive integers. For each $s \in [\![1, K]\!]$, there exist an encoding function $g_s : [\![1, M]\!] \to \mathcal{X}$ and a POVM $\mathbf{\Gamma}_s = \{\Gamma_s^w\}_{w \in [\![1,M]\!]}$ such that*

$$\frac{1}{MK} \sum_{s=1}^{K} \sum_{w=1}^{M} \text{tr}\left(\Gamma_s^w \rho_B^{g_s(w)}\right) \geqslant 1 - \epsilon, \text{ and } \left\| \frac{1}{MK} \sum_{s=1}^{K} \sum_{w=1}^{M} \rho_B^{g_s(w)} - \rho_B \right\|_1 \leqslant \delta, \tag{8.24}$$

provided that

$$\log M \leqslant \sup_{a \in]0,1[} \left[\aleph^a \left(X; B\right)_\rho - \frac{1}{a} \log \nu(\rho_X \otimes \rho_B) - \frac{4}{a} \log \frac{12}{\epsilon} \right], \tag{8.25}$$

and

$$\log MK \geqslant \inf_{a<0} \left[\beth^a \left(X;B \right)_\rho + \log \nu(\rho_B) + \left(2 - \frac{2}{a} \right) \log \frac{12}{\delta} \right]. \tag{8.26}$$

Proof. We consider K independently generated random encoders $G_s : [\![1, M]\!] \to \mathcal{X}$ where $G_s(1), \cdots, G_s(M)$ are iid according to P_X. By Lemma 4 in Chapter 1, for all $\epsilon > 0$, there exists a POVM such that

$$\mathbb{E}_{G_s} \left(\frac{1}{M} \sum_{w=1}^{M} \mathrm{tr} \left(\Gamma_s^w \rho_B^{G_s(w)} \right) \right) \geqslant 1 - \epsilon, \tag{8.27}$$

if $\log M \leqslant D_H^{\epsilon/2}(\rho_{XB} \| \rho_X \otimes \rho_B) - \log \frac{1}{\epsilon} - 4$ where

$$D_H^\epsilon(\rho \| \sigma) \triangleq - \log \inf_{Q : 0 \leqslant Q \leqslant \mathbf{1}, \mathrm{tr}(Q\rho) \geqslant 1-\epsilon} \mathrm{tr}\left(Q\sigma \right). \tag{8.28}$$

By [121, Theorem 2], for all $\gamma > 0$, there exists an operator $0 \leqslant Q \leqslant \mathbf{1}_{XB}$ such that for all $a \in]0, 1[$,

$$\mathrm{tr}\left(Q(\rho_X \otimes \rho_B) \right) \leqslant \nu(\rho_X \otimes \rho_B) e^{-\aleph^a(X;B)_\rho a - \gamma(1-a)} \tag{8.29}$$

$$\mathrm{tr}\left((\mathbf{1}_{XB} - Q)\rho_{XB} \right) \leqslant \nu(\rho_X \otimes \rho_B) e^{-\aleph^a(X;B)_\rho a + \gamma a}. \tag{8.30}$$

Choosing $\gamma = \aleph^a \left(X;B \right)_\rho - a^{-1}(\log \nu(\rho_X \otimes \rho_B) + \log 2/\epsilon)$ yields that

$$D_H^{\epsilon/2}(\rho_{XB} \| \rho_X \otimes \rho_B) - \log \frac{1}{\epsilon} - 4 \tag{8.31}$$

$$\geqslant \aleph^a \left(X;B \right)_\rho - \frac{1}{a} \log \nu(\rho_X \otimes \rho_B) - \frac{1}{a} \log \frac{2}{\epsilon} - \log \frac{1}{\epsilon} - 4 \tag{8.32}$$

$$\geqslant \aleph^a \left(X;B \right)_\rho - \frac{1}{a} \log \nu(\rho_X \otimes \rho_B) - \frac{4}{a} \log \frac{4}{\epsilon}. \tag{8.33}$$

To obtain (8.24), note that by Lemma 6 in Chapter 1, we have for all $a < 0$

$$\mathbb{E}_{G_1,\cdots,G_K}\left(\left\|\frac{1}{MK}\sum_{s=1}^{K}\sum_{w=1}^{M}\rho_B^{G_s(w)} - \rho_B\right\|_1\right) \leqslant \inf_{\lambda>0} 2\sqrt{e^{a(\log\lambda - \beth^a(X;B)_\rho)}} + \sqrt{\frac{\lambda\nu(\rho_B)}{MK}}. \tag{8.34}$$

Choosing $\log\lambda = \beth^a(X;B)_\rho + \frac{2}{a}\log\delta/4$ and $\log MK \geqslant -2\log\frac{\delta}{2} + \log\lambda + \log\nu(\rho_B)$, we obtain that

$$\mathbb{E}_{G_1,\cdots,G_K}\left(\left\|\frac{1}{MK}\sum_{s=1}^{K}\sum_{w=1}^{M}\rho_B^{G_s(w)} - \rho_B\right\|_1\right) \leqslant \delta. \tag{8.35}$$

Finally, Markov's inequality and the bounds on the expected values imply the existence of the desired code. □

We now prove the existence of a code for transmission of a classical message over a *noiseless* classical channel while a pre-specified distribution is induced at the output of the channel. We show that no key is required in this case. The idea of the proof is similar to [112, Lemma 2].

Lemma 59. *Let P_X be a PMF over $\mathcal{X}$, and Q_W be uniform distribution over $[\![1, M]\!]$ for $M \in \mathbb{N}^+$. Let $(W, X, \widehat{W})$ be distributed according to*

$$Q_{WX\widehat{W}}(w, x, \widehat{w}) \triangleq Q_W(w)Q_{X|W}(x|w)\mathbb{1}\{f(x) = \widehat{w}\} \tag{8.36}$$

for a conditional PMF $Q_{X|W}$ and a function $f : \mathcal{X} \to [\![1, M]\!]$. For all $\epsilon > 0$, there exists $Q_{X|W}$ and f such that

$$\|Q_X - P_X\|_1 \leqslant \epsilon, \tag{8.37}$$

$$\mathbb{P}_Q\left(W \neq \widehat{W}\right) \leqslant \epsilon, \tag{8.38}$$

provided that $\log M \leqslant \sup_{a\in]0,1[} H^a(P_X) - \frac{4}{a}\log\frac{2}{\epsilon}$.

Proof. Let $P_{WX\widehat{W}}$ be another distribution for $(W, X, \widehat{W})$ defined as

$$P_{WX\widehat{W}}(w, x, \widehat{w}) = P_X(x)\mathbb{1}\{g(x) = w, g(x) = \widehat{w}\}$$

for a function $g : \mathcal{X} \to [\![1, M]\!]$. Using a privacy amplification result [30, Corollary 5.6.1] and a bound on smooth min-entropy in terms of Rényi entropy [122, Theorem 7], there exists g such that $\|P_W - Q_W\|_1 \leqslant \epsilon$ when $\log M \leqslant \sup_{a\in]0,1[} H^a(P_X) - \frac{4}{a}\log\frac{2}{\epsilon}$. It is enough to show that (8.37) and (8.38) hold for $f \triangleq g$ and

$$Q_{X|W}(x|w) \triangleq \begin{cases} \frac{P_{XW}(x,w)}{P_W(w)} & P_W(w) \neq 0 \\ P_X(x) & P_W(w) = 0 \end{cases} \tag{8.39}$$

Note that $P_{\widehat{W}|XW}(\widehat{w}|x, w) = \mathbb{1}\{f(x) = \widehat{w}\}$ and we have $P_{XW}(x, w) = P_W(w)Q_{X|W}(x|w)$ for all w, x. We thus have

$$\begin{aligned} &\left\|Q_{WX\widehat{W}} - P_{WX\widehat{W}}\right\|_1 && (8.40)\\ &= \sum_{x,w,\widehat{w}} |Q_W(w)Q_{X|W}(x|w)\mathbb{1}\{f(x) = \widehat{w}\} - P_W(w)Q_{X|W}(x|w)\mathbb{1}\{f(x) = \widehat{w}\}\,| && (8.41)\\ &\leqslant \sum_w |Q_W(w) - P_W(w)| = \|Q_W - P_W\|_1 \leqslant \epsilon. && (8.42) \end{aligned}$$

By the data processing inequality, (8.37) holds. Since for any two distributions P and Q, we have $\|P - Q\|_1 = 2\sup_{\mathcal{A}} P(\mathcal{A}) - Q(\mathcal{A})$, we have

$$\begin{aligned} \mathbb{P}_Q\left(W \neq \widehat{W}\right) &\leqslant \mathbb{P}_P\left(W \neq \widehat{W}\right) + \frac{1}{2}\left\|Q_{WX\widehat{W}} - P_{WX\widehat{W}}\right\| && (8.43)\\ &= \frac{1}{2}\left\|Q_{WX\widehat{W}} - P_{WX\widehat{W}}\right\| \leqslant \epsilon. && (8.44) \end{aligned}$$

□

We extend Lemma 59 to the quantum setting in the following corollary.

Corollary 3. *Let $M \in \mathbb{N}^+$, $\epsilon > 0$, $\mathcal{H}$ be a finite dimensional Hilbert space, and ρ be a density operator on $\mathcal{H}$. Suppose that $\log M \leqslant \sup_{a\in]0,1[} H^a(\rho) - \frac{4}{a}\log\frac{2}{\epsilon}$. There exist a function $g : [\![1, M]\!] \to \mathcal{D}(\mathcal{H})$ and a POVM $\mathbf{\Lambda} = \{\Lambda^w\}_{w\in[\![1,M]\!]}$ such that*

$$\left\| \frac{1}{M}\sum_{w=1}^{M} g(w) - \rho \right\|_1 \leqslant \epsilon, \tag{8.45}$$

$$\frac{1}{M}\sum_{w=1}^{M} \mathrm{tr}\left(\Lambda^w g(w)\right) \geqslant 1 - \epsilon. \tag{8.46}$$

Proof. Considering an eigen-decomposition of ρ as $\sum_{i=1}^{d} P_X(x_i)|x_i\rangle\langle x_i|$ and defining $\mathcal{X} \triangleq \{x_1, \cdots, x_d\}$, we apply Lemma 59 to P_X to obtain a conditional PMF $Q_{X|W}$ and a function $f : \mathcal{X} \to [\![1, M]\!]$ satisfying (8.37) and (8.38). Let $Q_{WX\widehat{W}}$ be as defined in Lemma 59. We then define

$$g(w) \triangleq \sum_{x\in\mathcal{X}} Q_{X|W}(x|w)|x\rangle\langle x|, \tag{8.47}$$

$$\Lambda^w \triangleq \sum_{x:f(x)=w} |x\rangle\langle x|. \tag{8.48}$$

Substituting (8.47) in (8.45), we obtain

$$\left\| \frac{1}{M}\sum_{w=1}^{M} g(w) - \rho \right\|_1 = \left\| \frac{1}{M}\sum_{w=1}^{M}\sum_{x\in\mathcal{X}} Q_{X|W}(x|w)|x\rangle\langle x| - \rho \right\|_1 \tag{8.49}$$

$$= \left\| \sum_{x\in\mathcal{X}}\sum_{w=1}^{M} Q_W(w)Q_{X|W}(x|w)|x\rangle\langle x| - \rho \right\|_1 \tag{8.50}$$

$$= \left\| \sum_{x\in\mathcal{X}} Q_X(x)|x\rangle\langle x| - \sum_{x\in\mathcal{X}} P_X(x)|x\rangle\langle x| \right\|_1 \tag{8.51}$$

$$= \|Q_X - P_X\|_1 \leqslant \epsilon. \tag{8.52}$$

Moreover,

$$\frac{1}{M}\sum_{w=1}^{M}\mathrm{tr}\left(\Lambda^w g(w)\right) = \frac{1}{M}\sum_{w=1}^{M}\mathrm{tr}\left(\left(\sum_{x:f(x)=w}|x\rangle\langle x|\right)\left(\sum_{x\in\mathcal{X}}Q_{X|W}(x|w)|x\rangle\langle x|\right)\right) \tag{8.53}$$

$$= \frac{1}{M}\sum_{w=1}^{M}\mathrm{tr}\left(\sum_{x:f(x)=w}Q_{X|W}(x|w)|x\rangle\langle x|\right) \tag{8.54}$$

$$= \sum_{x,w}Q_W(w)Q_{X|W}(x|w)\mathbb{1}\{f(x)=w\} \tag{8.55}$$

$$= \sum_{x,w,\widehat{w}}Q_W(w)Q_{X|W}(x|w)\mathbb{1}\{f(x)=\widehat{w}\}\,\mathbb{1}\{w=\widehat{w}\} \tag{8.56}$$

$$= \mathbb{P}_Q\Big(W=\widehat{W}\Big) \leqslant \epsilon. \tag{8.57}$$

□

8.7.2 Proof of Main Results

Proof of Theorem 32. We separately prove the two parts of the theorem. Let the code $(f, \mathbf{\Lambda})$ be the cover protocol, and the main channel be noiseless. By Corollary 3, for every $\zeta \geqslant 0$, provided that

$$\log \overline{M} \leqslant \inf_{w}\sup_{a\in]0,1[} H^a(\mathcal{M}_{A\to A}^{\otimes n}(f(w))) - \frac{4}{a}\log\frac{2}{\epsilon}, \tag{8.58}$$

there exist a function $g_w : [\![1,\overline{M}]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n}\right)$ and a POVM $\mathbf{\Gamma}_w = \{\Gamma_w^{\overline{w}}\}_{\overline{w}\in[\![1,\overline{M}]\!]}$ such that

$$\left\|\frac{1}{\overline{M}}\sum_{\overline{w}=1}^{\overline{M}}g_w(\overline{w}) - \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right\|_1 \leqslant \zeta, \text{ and } \frac{1}{\overline{M}}\sum_{\overline{w}=1}^{\overline{M}}\mathrm{tr}\left(\Gamma_w^{\overline{w}}g_w(\overline{w})\right) \geqslant 1-\zeta. \tag{8.59}$$

We define the stego protocol as follows. Let $\overline{f} : [\![1, M\overline{M}]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n}\right)$ be defined as $\overline{f}(w(\overline{M}-1)+\overline{w}) \triangleq g_w(\overline{w})$. We define a POVM $\{\sqrt{\Lambda^w}\Gamma_w^{\overline{w}}\sqrt{\Lambda^w}\}_{w\in[\![1,M]\!],\overline{w}\in[\![1,\overline{M}]\!]}$, which is equivalent to first measuring $\mathbf{\Lambda}$ and then measuring $\mathbf{\Gamma}_w$. This is a valid POVM, since every $\sqrt{\Lambda^w}\Gamma_w^{\overline{w}}\sqrt{\Lambda^w}$ is a positive operator and $\sum_{w\overline{w}}\sqrt{\Lambda^w}\Gamma_w^{\overline{w}}\sqrt{\Lambda^w} \overset{(a)}{=} \sum_w \Lambda^w \overset{(b)}{=} \mathbf{1}_B$, where (a)

follows since $\mathbf{\Gamma}_w$ is a valid POVM, and (b) follows since $\mathbf{\Lambda}$ is a valid POVM. Note next that

$$\left\| \rho^c_{B^n} - \rho^s_{B^n} \right\|_1 = \left\| \frac{1}{M} \sum_{w=1}^{M} \mathcal{M}^{\otimes n}_{A\to A}(f(w)) - \frac{1}{M\overline{M}} \sum_{w\overline{w}} g_w(\overline{w}) \right\|_1 \tag{8.60}$$

$$\overset{(a)}{\leqslant} \frac{1}{M} \sum_{w}^{M} \left\| \mathcal{M}^{\otimes n}_{A\to A}(f(w)) - \frac{1}{\overline{M}} \sum_{\overline{w}} g_w(\overline{w}) \right\|_1 \leqslant \zeta, \tag{8.61}$$

where (a) follows from the convexity of the trace norm. The probability of correct decoding is also

$$\frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\sqrt{\Lambda^w} \Gamma^{\overline{w}}_w \sqrt{\Lambda^w} g_w(\overline{w}) \right) = \frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\Gamma^{\overline{w}}_w \sqrt{\Lambda^w} g_w(\overline{w}) \sqrt{\Lambda^w} \right) \tag{8.62}$$

$$= \frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\Gamma^{\overline{w}}_w g_w(\overline{w}) \right) + \frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\Gamma^{\overline{w}}_w \left(\sqrt{\Lambda^w} g_w(\overline{w}) \sqrt{\Lambda^w} - g_w(\overline{w}) \right) \right). \tag{8.63}$$

We also have $\frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\Gamma^{\overline{w}}_w g_w(\overline{w}) \right) \geqslant 1 - \zeta$ by (8.59). To lower-bound the second term in (8.63), we have

$$\frac{1}{M\overline{M}} \sum_{w\overline{w}} \text{tr}\left(\Gamma^{\overline{w}}_w \left(\sqrt{\Lambda^w} g_w(\overline{w}) \sqrt{\Lambda^w} - g_w(\overline{w}) \right) \right) \tag{8.64}$$

$$\geqslant -\frac{1}{M\overline{M}} \sum_{w\overline{w}} \left\| \Gamma^{\overline{w}}_w \right\|_\infty \left\| \sqrt{\Lambda^w} g_w(\overline{w}) \sqrt{\Lambda^w} - g_w(\overline{w}) \right\|_1 \tag{8.65}$$

$$\geqslant -\frac{1}{M\overline{M}} \sum_{w\overline{w}} \left\| \sqrt{\Lambda^w} g_w(\overline{w}) \sqrt{\Lambda^w} - g_w(\overline{w}) \right\|_1 \tag{8.66}$$

$$\overset{(a)}{\geqslant} -\frac{1}{M\overline{M}} \sum_{w\overline{w}} 2\sqrt{1 - \text{tr}\left(\Lambda^w g_w(\overline{w}) \right)} \tag{8.67}$$

$$\overset{(b)}{\geqslant} -2\sqrt{1 - \frac{1}{M} \sum_{w} \text{tr}\left(\Lambda^w \left(\frac{1}{\overline{M}} \sum_{\overline{w}} g_w(\overline{w}) \right) \right)}, \tag{8.68}$$

where (a) follows from the gentle operator lemma [123], and (b) follows from Jensen's inequality and the concavity of $x \mapsto \sqrt{1-x}$. We also lower-bound

$$\frac{1}{M}\sum_w \operatorname{tr}\left(\Lambda^w\left(\frac{1}{\overline{M}}\sum_{\overline{w}} g_w(\overline{w})\right)\right) \tag{8.69}$$

$$= \frac{1}{M}\sum_w \operatorname{tr}\left(\Lambda^w \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right) + \frac{1}{M}\sum_w \operatorname{tr}\left(\Lambda^w\left(\frac{1}{\overline{M}}\sum_{\overline{w}} g_w(\overline{w}) - \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right)\right) \tag{8.70}$$

$$\geqslant 1-\epsilon+\frac{1}{M}\sum_w \operatorname{tr}\left(\Lambda^w\left(\frac{1}{\overline{M}}\sum_{\overline{w}} g_w(\overline{w}) - \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right)\right) \tag{8.71}$$

$$\geqslant 1-\epsilon-\frac{1}{M}\sum_w \|\Lambda^w\|_\infty \left\|\frac{1}{\overline{M}}\sum_{\overline{w}} g_w(\overline{w}) - \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right\|_1 \tag{8.72}$$

$$\geqslant 1-\epsilon-\frac{1}{M}\sum_w \left\|\frac{1}{\overline{M}}\sum_{\overline{w}} g_w(\overline{w}) - \mathcal{M}_{A\to A}^{\otimes n}(f(w))\right\|_1 \geqslant 1-\epsilon-\zeta. \tag{8.73}$$

Let ζ, ξ, $\rho_{B^n}^w$, $\overline{M}_w$, $\overline{M}$, and $\overline{K}$ be defined as in the statement of Theorem 32, and the main channel be noisy. We assume without loss of generality that $\overline{M}_w$ is divisible by $\overline{M}$ for all w, otherwise we define $\overline{M}'_w \triangleq \lfloor \overline{M}_w/\overline{M} \rfloor \overline{M}$, and $|\log \overline{M}_w - \log \overline{M}'_w| \leqslant 1$. By Lemma 58, for each w, there exist $\overline{K}$ encoding functions $g_{s,w} : [\![1, \overline{M}_w]\!] \to \mathcal{D}(\mathcal{H}_A^{\otimes n})$ and $\overline{K}$ POVMs $\mathbf{\Gamma}_{s,w} = \{\Gamma_{s,w}^{\overline{w}}\}_{\overline{w}\in[\![1,\overline{M}_w]\!]}$ such that $\frac{1}{\overline{M}_w \overline{K}}\sum_{s\overline{w}} \operatorname{tr}\left(\Gamma_{s,w}^{\overline{w}}\rho_{B^n}^w\right) \geqslant 1-\zeta$, and

$$\left\|\frac{1}{\overline{M}_w \overline{K}}\sum_{s\overline{w}} \mathcal{N}_{A\to B}^{\otimes n}(g_{s,w}(\overline{w})) - \rho_{B^n}^w\right\|_1 \leqslant \xi. \tag{8.74}$$

We define the stego protocol as follows. For $s \in [\![1, K]\!]$, $w \in [\![1, M]\!]$, and $\overline{w} \in [\![1, \overline{M}]\!]$, we define $\mu_w \triangleq \frac{\overline{M}_w}{\overline{M}}$, $\overline{f}_s((w-1)\overline{M}+\overline{w}) \triangleq \frac{1}{\mu_w}\sum_{i=1}^{\mu_w} g_{s,w}((\overline{w}-1)\mu_w+i)$, and $\overline{\Lambda}_s^{(w-1)\overline{M}+\overline{w}} \triangleq \sum_{j=1}^{\mu_w} \sqrt{\Lambda^w}\Gamma_{s,w}^{(\overline{w}-1)\mu_w+j}\sqrt{\Lambda^w}$. As done previously, one can show that for each $s \in [\![1, K]\!]$, $\overline{\mathbf{\Lambda}}_s = \{\overline{\Lambda}_s\}$ is a valid POVM. Note also that

$$\frac{1}{K M \overline{M}}\sum_{s,w,\overline{w}} \operatorname{tr}\left(\Lambda_s^{(w-1)\overline{M}+\overline{w}} f_s((w-1)\overline{M}+\overline{w})\right) \tag{8.75}$$

$$= \frac{1}{KM\overline{M}} \sum_{s,w,\overline{w}} \text{tr} \left(\left(\sum_{i=1}^{\mu_w} \sqrt{\Lambda^w} \Gamma_{s,w}^{(\overline{w}-1)\mu_w+i} \sqrt{\Lambda^w} \right) \left(\frac{1}{\mu_w} \sum_{j=1}^{\mu_w} g_{s,w}((\overline{w}-1)\mu_w + j) \right) \right) \tag{8.76}$$

$$\geqslant \frac{1}{KM\overline{M}} \sum_{s,w,\overline{w}} \text{tr} \left(\frac{1}{\mu_w} \sum_{i=1}^{\mu_w} \sqrt{\Lambda^w} \Gamma_{s,w}^{(\overline{w}-1)\mu_w+i} \sqrt{\Lambda^w} g_{s,w}((\overline{w}-1)\mu_w + i) \right) \tag{8.77}$$

$$\overset{(a)}{=} \frac{1}{KM} \sum_{s,w} \frac{1}{\overline{M}_w} \sum_{\overline{w}=1}^{\overline{M}_w} \text{tr} \left(\sqrt{\Lambda^w} \Gamma_{s,w}^{\overline{w}} \sqrt{\Lambda^w} g_{s,w}((\overline{w}) \right), \tag{8.78}$$

where (a) follows since the index $(\overline{w}-1)\mu_w + i$ is changing from 1 to $\overline{M}_w$. Repeating calculations similar to (8.62)-(8.73), we obtain that

$$\frac{1}{KM\overline{M}} \sum_{s,w,\overline{w}} \text{tr} \left(\Lambda_s^{(w-1)\overline{M}+\overline{w}} f_s((w-1)\overline{M} + \overline{w}) \right) \geqslant 1 - \zeta - 2\sqrt{\zeta + \epsilon}. \tag{8.79}$$

Furthermore, we have

$$\left\| \frac{1}{M} \sum_{w=1}^{M} \rho_{B^n}^w - \frac{1}{\overline{K}M\overline{M}} \sum_{s,w\overline{w}} \mathcal{N}_{A\to B}^{\otimes n}(\overline{f}_s((w-1)\overline{M} + \overline{w})) \right\|_1 \tag{8.80}$$

$$\overset{(a)}{\leqslant} \frac{1}{M} \sum_{w=1}^{M} \left\| \rho_{B^n}^w - \frac{1}{\overline{KM}} \sum_{s\overline{w}} \mathcal{N}_{A\to B}^{\otimes n}(\overline{f}_s((w-1)\overline{M} + \overline{w})) \right\|_1 \tag{8.81}$$

$$= \frac{1}{M} \sum_{w=1}^{M} \left\| \rho_{B^n}^w - \frac{1}{\overline{KM}} \sum_{s\overline{w}} \mathcal{N}_{A\to B}^{\otimes n} \left(\frac{1}{\mu_w} \sum_{i=1}^{\mu_w} g_{s,w}((\overline{w}-1)\mu_w + i) \right) \right\|_1 \tag{8.82}$$

$$= \frac{1}{M} \sum_{w=1}^{M} \left\| \rho_{B^n}^w - \frac{1}{\overline{KM}_w} \sum_{s\overline{w}} \mathcal{N}_{A\to B}^{\otimes n} \left(g_{s,w}(\overline{w}) \right) \right\|_1 \leqslant \xi, \tag{8.83}$$

where (a) follows from the convexity of the trace norm.

□

Proof of Theorem 33. Intuitively, Alice splits the transmission into two part. Alice generates a purification of the state supposed to be transmitted in the first part, keeps the reference system for herself, and transmits the state over the channel, which results in a shared entan-

gled state between Alice and Bob. Alice and Bob use an entanglement distillation protocol to distill perfect entanglement in the second part of the transmission. This might require classical communication, which can be achieved by using the result of Theorem 32. To formally state our protocol, we first need a generalization of the gentle measurement lemma.

Proposition 7. *Suppose that $\rho^x \in \mathcal{D}(\mathcal{H})$ is a density operator, $\mathcal{N}^x : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H}')$ is a quantum channel for all $x \in \mathcal{X}$, and $\mathbf{\Lambda} = \{\Lambda^x\}_{x\in\mathcal{X}}$ is a POVM. Suppose that P_X is a PMF over $\mathcal{X}$ such that $\sum_x P_X(x)\mathrm{tr}\left(\rho^x\Lambda^x\right) \geqslant 1-\epsilon$. It then holds that*

$$\left\| \sum_x P_X(x) \left(\mathcal{N}^x(\rho^x) - \sum_{x'} \mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}}) \right) \right\|_1 \leqslant 2\sqrt{\epsilon} + \epsilon. \tag{8.84}$$

Proof. See Appendix 8.8. □

Let $(f, \mathbf{\Lambda})$ be the $(M,\epsilon)^{\mathrm{CC}}$ satisfying $f(w) = f_1(w) \otimes f_2(w)$ for all $w \in [\![1, M]\!]$, where $f_1 : [\![1, M]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n_1}\right)$, $f_2 : [\![1, M]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n_2}\right)$, and $n_1 + n_2 = n$. Let $\overline{M}^{\mathrm{CC}} \triangleq \overline{M}^{\mathrm{CC}}(f_2, \zeta)$ and $K \triangleq K^{\mathrm{CC}}(f_2, \zeta)$. Using the same argument as in the proof of Theorem 32, there exist an encoder function $g_w : [\![1, \overline{M}^{\mathrm{CC}}]\!] \times [\![1, K]\!] \to \mathcal{D}\left(\mathcal{H}_A^{\otimes n_2}\right)$ and a POVM $\mathbf{\Gamma}_{w,s} = \{\Gamma^{\overline{w}}_{w,s}\}_{\overline{w}\in[\![1,\overline{M}^{\mathrm{CC}}]\!]}$ for each message $w \in [\![1, M]\!]$ such that

$$\left\| \frac{1}{\overline{M}^{\mathrm{CC}} K} \sum_{\overline{w},s} \mathcal{N}_{A\to B}^{\otimes n_2}\left(g_w(\overline{w}, s)\right) - \left(\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A}\right)^{\otimes n_2}(f_2(w)) \right\|_1 \leqslant \zeta, \tag{8.85}$$

$$\frac{1}{\overline{M}^{\mathrm{CC}} K} \sum_{s\overline{w}} \mathrm{tr}\left(\Gamma^{\overline{w}}_{w,s} \mathcal{N}_{A\to B}^{\otimes n_2}(g_w(\overline{w}, s))\right) \geqslant 1 - \zeta. \tag{8.86}$$

Let $|\phi^w\rangle_{RA^{n_1}}$ and $\sigma^w_{RB^{n_1}}$ be defined as in Theorem 33. We define $\overline{M}^w \triangleq E_d(\rho^w_{RB^{n_1}}, \overline{M}^{\mathrm{CC}}, \zeta)$ and fix an $(\overline{M}^w, \overline{M}^{\mathrm{CC}}, \rho^w_{RB^{n_1}}, \zeta)^{\mathrm{ED}}$ protocol $(\mathcal{E}^w_{R\to C\widetilde{A}}, \mathcal{D}^w_{CB^{n_1}\to\widetilde{B}})$. Let S_1 and S_2 be two shared secret keys between Alice and Bob uniformly distributed over $[\![1, \overline{M}^{\mathrm{CC}}]\!]$ and $[\![1, K]\!]$, respectively. We define a POVM $\overline{\mathbf{\Lambda}}_{s_2} = \{\overline{\Lambda}^{w,\overline{w}}_{s_2}\}_{w,\overline{w}}$ with $\overline{\Lambda}^{w,\overline{w}}_{s_2} \triangleq \sqrt{\Lambda^w}(\mathbf{1}_{B^{n_1}} \otimes \Gamma^{\overline{w}}_{w,s_2})\sqrt{\Lambda^w}$. The stego protocol would operate as follows when $W = w$.

Alice prepares $|\phi^w\rangle_{RA^{n_1}}$ and sends $\phi^w_{A^{n_1}}$ over n_1 uses of $\mathcal{N}_{A\to B}$. Alice then applies

$\mathcal{E}^w_{R\to C\widetilde{A}}$ to ϕ^w_R and sends $g_w(C \oplus S_1, S_2)$ over n_2 uses of $\mathcal{N}_{A\to B}$. Bob performs the POVM $\overline{\mathbf{\Lambda}}_{S_2}$ to decode W and C with the help of S_1. Bob finally applies $\mathcal{D}^w_{CB^{n_1}\to\widetilde{B}}$ to his first n_1 received subsystem to obtain the entangled state.

Let $\rho^w_{B^n}$ denote the state received by Bob when $W = w$ and the cover protocol is executed over n uses of $\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to B}$. Let $\overline{\rho}^w_{B^n}$ denote the state received by Bob when $W = w$ and the stego protocol is executed over n uses of $\mathcal{N}_{A\to B}$. Note that both $\rho^w_{B^n}$ and $\overline{\rho}^w_{B^n}$ decompose as

$$\rho^w_{B^n} = (\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A})^{\otimes n_1}(f_1(w)) \otimes (\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A})^{\otimes n_2}(f_2(w)) \tag{8.87}$$

$$\overline{\rho}^w_{B^n} = \mathcal{N}^{\otimes n_1}_{A\to B}(\phi^w_{A^{n_1}}) \otimes \mathcal{N}^{\otimes n_2}_{A\to B}\left(\sum_{\overline{w},s_2} \mathbb{P}(C + S_1 = \overline{w}, S_2 = s_2) g_w(\overline{w}, s_2)\right) \tag{8.88}$$

$$= \mathcal{N}^{\otimes n_1}_{A\to B}(\phi^w_{A^{n_1}}) \otimes \left(\frac{1}{\overline{M}^{\mathrm{CC}} K} \sum_{\overline{w}s_2} \mathcal{N}^{\otimes n_2}_{A\to B}\left(g_w(\overline{w}, s_2)\right)\right). \tag{8.89}$$

We have $(\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A})^{\otimes n_1}(f_1(w)) = \mathcal{N}^{\otimes n_1}_{A\to B}(\phi^w_{A^{n_1}})$ because $|\phi^w\rangle_{RA^{n_1}}$ is a purification of $\mathcal{M}^{\otimes n_1}_{A\to A}(f_1(w))$. We therefore have

$$\left\|\rho^w_{B^n} - \overline{\rho}^w_{B^n}\right\|_1 = \left\|(\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A})^{\otimes n_2}(f_2(w)) - \frac{1}{\overline{M}^{\mathrm{CC}} K} \sum_{\overline{w}s_2} \mathcal{N}^{\otimes n_2}_{A\to B}\left(g_w(\overline{w}, s_2)\right)\right\|_1, \tag{8.90}$$

which is less than ζ by (8.85). By the convexity of trace norm, it holds that $\left\|\rho^c_{B^n} - \rho^s_{B^n}\right\|_1 \leqslant \zeta$. Following the same reasoning of the proof of Theorem 32, we conclude that

$$\frac{1}{M\overline{M}^{\mathrm{CC}} K} \frac{1}{\overline{M}^{\mathrm{CC}}, M} \sum_{w,\overline{w},,s_2} \mathrm{tr}\left(\overline{\Lambda}^{w,\overline{w}}_{s_1} \mathcal{N}^{\otimes n}_{A\to B}(\phi^w_{A^{n_1}} \otimes g_w(\overline{w}, s_2))\right) \geqslant 1 - \zeta - 2\sqrt{\epsilon + \zeta}. \tag{8.91}$$

In other words, Bob correctly decodes W and C with probability at least $1 - \zeta - 2\sqrt{\epsilon + \zeta}$.

We fix $W = w$, $S_1 = s_1$, and $S_2 = s_2$ and denote

$$(\mathcal{E}^w_{R\to C\widetilde{A}} \otimes \mathrm{id}_{A^{n_1}})(\phi^w_{RA^{n_1}}) = \sum_c \mathbb{P}(C = c|W = w)|c\rangle\langle c|_C \otimes \phi^{w,c}_{\widetilde{A}A^{n_1}}. \tag{8.92}$$

Fixing a value $C = c$ and setting $\overline{w} \triangleq c \oplus s_1$, Alice transmits the subsystem A^n of $\phi^{w,c}_{\widetilde{A}A^{n_1}} \otimes g_w(\overline{w}, s_2)$ over $\mathcal{N}^{\otimes n}_{A\to B}$, which results in the state $\phi^{w,c}_{\widetilde{A}B^n}$.

The shared entangled state would be

$$\sum_{w'\overline{w}'} \mathrm{id}_{\widetilde{A}} \otimes \mathcal{D}^{w'}_{CB^{n_1}\to\widetilde{B}} \otimes \emptyset_{B^{n_2}} \left(|\overline{w}' - s_1\rangle\langle\overline{w}' - s_1| \otimes (\mathbf{1}_{\widetilde{A}} \otimes \sqrt{\overline{\Lambda}^{w'\overline{w}'}_{s_2}})\phi^{w,c}_{\widetilde{A}B^n}(\mathbf{1}_{\widetilde{A}} \otimes \sqrt{\overline{\Lambda}^{w'\overline{w}'}_{s_2}}) \right) \tag{8.93}$$

By (8.91) and Proposition 7, we obtain that

$$\begin{aligned}
\Bigg\| \sum_{wcs_2} \mathbb{P}(W = w, C = c, S_2 = s_2) \sum_{w'\overline{w}'} \mathrm{id}_{\widetilde{A}} \otimes \mathcal{D}^{w'}_{CB^{n_1}\to\widetilde{B}} \otimes \emptyset_{B^{n_2}} \\
\left(|\overline{w}' - s_1\rangle\langle\overline{w}' - s_1| \otimes (\mathbf{1}_{\widetilde{A}} \otimes \sqrt{\overline{\Lambda}^{w'\overline{w}'}_{s_2}})\phi^{w,c}_{\widetilde{A}B^n}(\mathbf{1}_{\widetilde{A}} \otimes \sqrt{\overline{\Lambda}^{w'\overline{w}'}_{s_2}}) \right) \\
- \sum_{wcs_2} \mathbb{P}(W = w, C = c, S_2 = s_2) \mathrm{id}_{\widetilde{A}} \otimes \mathcal{D}^{w}_{CB^{n_1}\to\widetilde{B}} \otimes \emptyset_{B^{n_2}} \left(|c\rangle\langle c| \otimes \phi^{w,c}_{\widetilde{A}B^n} \right) \Bigg\|_1 \\
\leqslant \zeta + 2\sqrt{\epsilon + \zeta} + 2\sqrt{\zeta + 2\sqrt{\epsilon + \zeta}}.
\end{aligned} \tag{8.94}$$

By the definition of an entanglement distillation code, we have

$$\left\| \sum_c \mathbb{P}(C = c|W = w) \mathrm{id}_{\widetilde{A}} \otimes \mathcal{D}_{CB^{n_1}\to\widetilde{B}}(|c\rangle\langle c|_C \otimes \phi^{w,c}_{\widetilde{A}B^{n_1}}) - \Phi^{(\overline{M})} \right\|_1 \leqslant \zeta. \tag{8.95}$$

Using the triangle inequality completes the proof.

□

Proof of Theorem 34. Let $|\phi\rangle_{R\widetilde{A}A^n}$ be a purification of $\rho_{\widetilde{A}A^n}$. Let $V_{A\to BE}$ and $W_{B^n\to\widetilde{B}H}$ be isometric extensions of $\mathcal{N}_{A\to B} \circ \mathcal{M}_{A\to A} = \mathcal{M}_{A\to A}$ and $\mathcal{D}_{B^n\to\widetilde{B}}$, respectively. The stego

protocol will be as follows. Alice prepares a pure state $|\omega\rangle_{R\widetilde{A}B^nE^n} \triangleq \mathbf{1}_{R\widetilde{A}} \otimes V^{\otimes n}_{A\to BE}|\phi\rangle_{R\widetilde{A}A^n}$ and sends ω_{B^n} over $\mathcal{N}^{\otimes n}_{A\to B}$. Bob applies $W_{B^n\to\widetilde{B}H}$ on ω_{B^n}, which results in the overall state

$$|\psi\rangle_{R\widetilde{A}E^n\widetilde{B}H} \triangleq (\mathbf{1}_{R\widetilde{A}E^n} \otimes W_{B^n\to\widetilde{B}H}) \circ (\mathbf{1}_{R\widetilde{A}} \otimes V^{\otimes n}_{A\to BE})|\phi\rangle_{R\widetilde{A}A^n}, \tag{8.96}$$

Note that $F(\Phi^{(M)}, \psi_{\widetilde{A}\widetilde{B}}) \geqslant 1-\epsilon$ from our assumption on the code. We now follow a standard application of Uhlmann's theorem to show that Bob can indeed decode ψ_{RE}. Note that $|\psi\rangle_{R\widetilde{A}E^n\widetilde{B}H}$ is a purification of $\psi_{\widetilde{A}\widetilde{B}}$. The state $\Phi^{(M)}$ also has a purification over $\mathcal{H}_{R\widetilde{A}E^n\widetilde{B}H}$. Uhlmann's theorem therefore implies the existence of a purification $|\tau\rangle_{R\widetilde{A}E^n\widetilde{B}H}$ of $\Phi^{(M)}$ over $\mathcal{H}_{R\widetilde{A}E^n\widetilde{B}H}$ such that $|\langle\tau|\psi\rangle_{R\widetilde{A}E^n\widetilde{B}H}|^2 = F(\Phi^{(M)}, \psi_{\widetilde{A}\widetilde{B}})$. The vector $|\Phi^{(M)}\rangle \otimes |0\rangle$ is another purification of $\Phi^{(M)}$ for every unit vector $|0\rangle \in \mathcal{H}_{RE^nH}$. By [85], there exists a unitary $T_{RE^nH\to RE^nH}$ on $\mathcal{H}_{RE^nH}$ such that

$$\begin{aligned}
|\tau\rangle_{R\widetilde{A}E^n\widetilde{B}H} &= (\mathbf{1}_{\widetilde{A}\widetilde{B}} \otimes T_{RE^nH\to RE^nH})|\Phi^{(M)}\rangle_{\widetilde{A}\widetilde{B}} \otimes |0\rangle && (8.97)\\
&= |\Phi^{(M)}\rangle \otimes (T_{RE^nH\to RE^nF^nH}|0\rangle) && (8.98)\\
&\triangleq |\Phi^{(M)}\rangle \otimes |\tau'\rangle_{RE^nH}. && (8.99)
\end{aligned}$$

We thus have $\langle\tau|\psi\rangle_{R\widetilde{A}E^n\widetilde{B}H} = (\langle\Phi^{(M)}| \otimes \langle\tau|_{RE^nH})|\psi\rangle_{R\widetilde{A}E^n\widetilde{B}H}$. We consider a Schmidt decomposition of $|\tau'\rangle_{RE^nH}$ such as $|\tau'\rangle_{RE^nH} = \sum_{x\in\mathcal{X}} \sqrt{P_X(x)}|\alpha_x\rangle_{RE^n} \otimes |\beta_x\rangle_H$, where P_X is a PMF over $\mathcal{X}$, and $|\alpha_x\rangle_{RE^n}$ and $|\beta_x\rangle_H$ are orthonormal in RE^n and H, respectively. Let X be a random variable distributed according to P_X, $f : \mathcal{X} \to [\![1, \overline{M}]\!]$, and Q and $P_{f(X)}$ be the uniform distribution and the distribution of $f(X)$, respectively. By [30, Corollary 5.6.1] and [122, Theorem 7], when $\log \overline{M} \leqslant \sup_{a\in]0,1[} H^a(P_X) - \frac{4}{a}\log\frac{2}{\zeta}$, there exists a function f such that $\left\|Q - P_{f(X)}\right\|_1 \leqslant \zeta$. Alice measures $\{\Lambda^{\overline{w}}_A \triangleq \sum_{x:f(x)=\overline{w}} |\alpha_x\rangle\langle\alpha_x|_{RE^n}\}_{\overline{w}\in[\![1,\overline{M}]\!]}$ on ψ_{RE^n} and Bob measures $\{\Lambda^{\overline{w}}_B \triangleq \sum_{x:f(x)=\overline{w}} |\beta_x\rangle\langle\beta_x|_H\}_{\overline{w}\in[\![1,\overline{M}]\!]}$ on ψ_H. Let W_A and W_B denote the output of the Alice's and Bob's measurement, respectively and $\mathcal{A}_{RE^n\to W_A}$ and

$\mathcal{B}_{H\to W_B}$ denote the corresponding quantum channels to these measurements. We have

$$\sqrt{F((\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\psi_{\widetilde{A}\widetilde{B}RE^nH}),\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})})} \tag{8.100}$$

$$\geqslant 1-\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\psi_{\widetilde{A}\widetilde{B}RE^nH})-\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})}\right\|_1 \tag{8.101}$$

$$\begin{aligned}\geqslant 1-\big\|&(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\psi_{\widetilde{A}\widetilde{B}RE^nH})\\&-(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})\big\|_1\end{aligned} \tag{8.102}$$

$$-\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})-\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})}\right\|_1 \tag{8.103}$$

$$\begin{aligned}\geqslant 1-&\left\|\psi_{\widetilde{A}\widetilde{B}RE^nH}-\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH}\right\|_1\\&-\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})-\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})}\right\|_1\end{aligned} \tag{8.104}$$

$$\begin{aligned}\geqslant 1-&\sqrt{1-F(\psi_{\widetilde{A}\widetilde{B}RE^nH},\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})}\\&-\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})-\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})}\right\|_1\end{aligned} \tag{8.105}$$

$$\geqslant 1-\sqrt{\epsilon}-\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH})-\Phi^{(M)}\otimes\overline{\Phi}^{(\overline{M})}\right\|_1. \tag{8.106}$$

We can also write

$$(\mathrm{id}_{\widetilde{A}\widetilde{B}}\otimes\mathcal{A}_{RE^n\to W_A}\otimes\mathcal{B}_{H\to W_B})(\Phi^{(M)}\otimes|\tau'\rangle\langle\tau'|_{RE^nH}) \tag{8.107}$$

$$=\Phi^{(M)}\otimes\left(\sum_x P_X(x)|f(x)f(x)\rangle\langle f(x)f(x)|_{W_AW_B}\right) \tag{8.108}$$

$$=\Phi^{(M)}\otimes\left(\sum_{\overline{w}} P_{f(X)}(\overline{w})|\overline{ww}\rangle\langle\overline{ww}|_{W_AW_B}\right). \tag{8.109}$$

Hence,

$$\left\|(\mathrm{id}_{\widetilde{A}\widetilde{B}} \otimes \mathcal{A}_{RE^n \to W_A} \otimes \mathcal{B}_{H \to W_B})(\Phi^{(M)} \otimes |\tau'\rangle\langle\tau'|_{RE^n H}) - \Phi^{(M)} \otimes \overline{\Phi}^{(\overline{M})}\right\|_1 \leqslant \left\|P_{f(W)} - Q\right\|_1 \leqslant \zeta. \quad (8.110)$$

□

Proof of Theorem 35. We start the proof by a technical lemma that helps us simplify the expression of the rate of the cypher message. Let $(\mathcal{E}_{W\to A}, \mathcal{D}_{A\to W})$ be an $(M,\epsilon)^{\mathrm{QC}}_{\mathrm{R}}$ code for one use of the channel $\mathcal{M}_{A\to A}$. Suppose that $\mathcal{E}_{W\to A}(\rho_W) = V_{W\to A}\rho_W V^\dagger_{W\to A}$ where $V_{W\to A}$ is an isometry, and $\Pi = V_{W\to A}V^\dagger_{W\to A}$ is the projector on to the range of $V_{W\to A}$. Consider a decomposition, $\mathcal{M}_{A\to A} = \widetilde{\mathcal{M}}_{A\to A} + \widetilde{\widetilde{\mathcal{M}}}_{A\to A}$ such that $\mathcal{D}_{B\to W} \circ \widetilde{\mathcal{M}}_{A\to A} \circ \mathcal{E}_{W\to A} = c\,\mathrm{id}_W$ for $c \geqslant 1-\epsilon$. There exists a Kraus representation $\{F_j\}_{j\in\mathcal{J}}$ for $\widetilde{\mathcal{M}}_{A\to A}$ such that $\Pi F_j^\dagger F_{j'}\Pi = \mathbb{1}\{j=j'\}\, d_j \Pi$ for real positive numbers $\{d_j : j \in \mathcal{J}\}$. Define a PMF P_J over $\mathcal{J}$ as $P_J(j) \triangleq \frac{d_j}{\sum_{j'} d_{j'}}$.

Lemma 60. *For all $\delta > 2\sqrt{\epsilon}$, we have $H^\delta_{\min}(P_J) \geqslant H^{\delta-2\sqrt{\epsilon}}_{\min}\left(\mathcal{M}^c_{A\to A}\left(\frac{1}{M}\Pi\right)\right)$, where $\mathcal{M}^c_{A\to A}$ is the complementary channel of $\mathcal{M}_{A\to A}$.*

Proof. See Appendix 8.9 □

Consider the $(M,\epsilon)^{\mathrm{QC}}_{\mathrm{R}}$ cover protocol $(\mathcal{E}_{W\to A^n}, \mathcal{D}_{B^n\to W})$ for the channel $\mathcal{M}^{\otimes n}_{A\to A}$. Let $\mathcal{E}_{W\to A^n}(\rho) = V_{W\to A^n}\rho V^\dagger_{W\to A^n}$ where $V_{W\to A^n}$ is an isometry, and Π denote the projector onto the range of $V_{W\to A^n}$. By definition, there exists a decomposition $\mathcal{M}^{\otimes n}_{A\to A} = \widetilde{\mathcal{M}}_{A^n\to A^n} + \widetilde{\widetilde{\mathcal{M}}}_{A^n\to A^n}$ such that $\mathcal{D}_{B^n\to W} \circ \widetilde{\mathcal{M}}_{A^n\to A^n} \circ \mathcal{E}_{W\to A^n} = c\,\mathrm{id}_W$ with $c \geqslant 1-\epsilon$. By the same argument as in the proof of [32, Theorem 10.1], there exists a Kraus representation $\{F_j\}_{j\in\mathcal{J}}$ for $\widetilde{\mathcal{M}}_{A^n\to A^n}$ such that $\Pi F_j^\dagger F_{j'}\Pi = \mathbb{1}\{j=j'\}\, d_j\Pi$. By polar decomposition, we therefore have $F_j\Pi \triangleq U_j\sqrt{\Pi F_j^\dagger F_j \Pi} = \sqrt{d_j}U_j\Pi$ for some unitary U_j on $\mathcal{H}^{\otimes n}_A$.

Let J be distributed according to $P_J(j) \triangleq \frac{d_j}{\sum_{j'} d_{j'}}$, and Q denote the uniform distribution over $[\![1,\overline{M}]\!]$. By [30, Corollary 5.6.1], there exists a function $g : \mathcal{J} \to [\![1,\overline{M}]\!]$ such that

$\left\|P_{g(J)} - Q\right\|_1 \leqslant \zeta$, provided that

$$\log \overline{M} = H_{\min}^{\zeta/2}(P_J) - 2\log\frac{2}{\zeta} \tag{8.111}$$

$$\overset{(a)}{\geqslant} H_{\min}^{\zeta/2-2\sqrt{\epsilon}}\left(\mathcal{M}_{A\to A}^{c}{}^{\otimes n}\left(\frac{1}{M}\Pi\right)\right) - 2\log\frac{2}{\zeta} \tag{8.112}$$

$$\overset{(b)}{\geqslant} \sup_{a\in]0,1[} H^a\left(\mathcal{M}_{A\to A}^{c}{}^{\otimes n}\left(\frac{1}{M}\Pi\right)\right) - 2\log\frac{2}{\epsilon} - \frac{1}{a}\log\frac{2}{(\zeta/2-2\sqrt{\epsilon})^2} \tag{8.113}$$

$$\geqslant \sup_{a\in]0,1[} H^a\left(\mathcal{M}_{A\to A}^{c}{}^{\otimes n}\left(\frac{1}{M}\Pi\right)\right) - \frac{4}{a}\log\frac{2}{\zeta/2-2\sqrt{\epsilon}}, \tag{8.114}$$

where (a) follows from Lemma 60, and (b) follows from [122, Theorem 7].

Let $\mu_{\overline{w}} \triangleq \sum_{j\in\widetilde{\mathcal{J}}} \mathbb{1}\{g(j) = \overline{w}\}$. We then define

$$\overline{\mathcal{E}}_{W\to A^n}^{\overline{w}}(\rho) \triangleq \frac{1}{\mu_{\overline{w}}} \sum_{j:g(j)=\overline{w}} U_j \mathcal{E}_{W\to A^n}(\rho) U_j^\dagger \tag{8.115}$$

(for $\mu_{\overline{w}} = 0$ take $\overline{\mathcal{E}}_{W\to A^n}^{\overline{w}} = \mathcal{E}_{W\to A^n}$). We define the decoder for Bob as

$$\begin{aligned}&\overline{\mathcal{D}}_{B^n\to W\overline{W}}(\rho_{B^n})\\ &\quad\triangleq (\mathcal{D}_{B^n\to W}\otimes \mathrm{id}_{\overline{W}})\left(\sum_j (PU_j^\dagger \otimes |g(j)\rangle)\rho_{B^n}(U_jP\otimes\langle g(j)|) + E\rho_{B^n}E^\dagger\right),\end{aligned} \tag{8.116}$$

where the term $E\rho_{B^n}^\dagger E^\dagger$ is added to ensure that $\overline{\mathcal{D}}_{B^n\to W\overline{W}}$ is trace-preserving. By the argument in the proof of [32, Theorem 10.1], $\overline{\mathcal{D}}_{B^n\to W\overline{W}}$ is a valid quantum channel. The partial channels are

$$\overline{\mathcal{D}}_{B^n\to W}(\rho_{B^n}) \triangleq \mathcal{D}_{B^n\to W}\left(\sum_j (PU_j^\dagger)\rho_{B^n}(U_jP) + E'\rho_{B^n}E'^\dagger\right), \tag{8.117}$$

$$\overline{\mathcal{D}}_{B^n\to\overline{W}}(\rho_{B^n}) \triangleq \sum_j \mathrm{tr}\left(PU_j^\dagger\rho_{B^n}U_jP\right)|g(j)\rangle\langle g(j)| + E''\rho_{B^n}E''^\dagger. \tag{8.118}$$

Furthermore, for any $\rho_W \in \mathcal{D}(\mathcal{H}_W)$, we have

$$\frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \operatorname{tr}\left(|\overline{w}\rangle\langle\overline{w}| \mathcal{D}_{B^n \to \overline{W}}(\mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W))\right) \tag{8.119}$$

$$= \frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \operatorname{tr}\left(|\overline{w}\rangle\langle\overline{w}| \left(\sum_j \operatorname{tr}\left(PU_j^\dagger \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) U_j P\right) |g(j)\rangle\langle g(j)| \right.\right.$$

$$\left.\left. + E'' \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) E''^\dagger\right)\right) \tag{8.120}$$

$$\geqslant \frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \operatorname{tr}\left(|\overline{w}\rangle\langle\overline{w}| \left(\sum_j \operatorname{tr}\left(PU_j^\dagger \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) U_j P\right) |g(j)\rangle\langle g(j)|\right)\right) \tag{8.121}$$

$$= \frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \sum_{j:g(j)=\overline{w}} \operatorname{tr}\left(PU_j^\dagger \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) U_j P\right) \tag{8.122}$$

$$\geqslant \frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \sum_{j:g(j)=\overline{w}} \mathbb{1}\{\mu_{\overline{w}} \neq 0\} \operatorname{tr}\left(PU_j^\dagger \left(\frac{1}{\mu_{\overline{w}}} \sum_{j':g(j')=\overline{w}} U_{j'} \mathcal{E}_{W\to A^n}(\rho) U_{j'}^\dagger\right) U_j P\right) \tag{8.123}$$

$$\geqslant \frac{1}{\overline{M}} \sum_{\overline{w}=1}^{\overline{M}} \frac{1}{\mu_{\overline{w}}} \sum_{j:g(j)=\overline{w}} \mathbb{1}\{\mu_{\overline{w}} \neq 0\} \operatorname{tr}\left(PU_j^\dagger U_j \mathcal{E}_{W\to A^n}(\rho) U_j^\dagger U_j P\right) \tag{8.124}$$

$$= \operatorname{tr}\left(P \mathcal{E}_{W\to A^n}(\rho_W) P\right) \frac{1}{\overline{M}} \sum_{\overline{w}} \mathbb{1}\{\mu_{\overline{w}} \neq 0\} \tag{8.125}$$

$$= \frac{1}{\overline{M}} \sum_{\overline{w}} \mathbb{1}\{\mu_{\overline{w}} \neq 0\} \geqslant 1 - \left\|P_{g(J)} - Q\right\|_1 \tag{8.126}$$

$$\geqslant 1 - \zeta.$$

For a $\overline{w} \in [\![1, \overline{M}]\!]$ we have

$$\overline{\mathcal{D}}_{B^n\to W} \circ \widetilde{\mathcal{M}}_{A^n\to A^n} \circ \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) \tag{8.127}$$

$$= \mathcal{D}_{B^n\to W}\left(\sum_j (PU_j^\dagger) \widetilde{\mathcal{M}}_{A^n\to A^n} \circ \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W)(U_j P) + E' \widetilde{\mathcal{M}}_{A^n\to A^n} \circ \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W) E'^\dagger\right) \tag{8.128}$$

$$= \mathcal{D}_{B^n\to W}\left(\sum_j (PU_j^\dagger) \widetilde{\mathcal{M}}_{A^n\to A^n} \circ \mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W)(U_j P)\right). \tag{8.129}$$

We can write

$$\sum_j (PU_j^\dagger)\widetilde{\mathcal{M}}_{A^n\to A^n}\circ\mathcal{E}^{\overline{w}}_{W\to A^n}(\rho_W)(U_jP) \tag{8.130}$$

$$= \sum_j (PU_j^\dagger)\widetilde{\mathcal{M}}_{A^n\to A^n}\left(\frac{1}{\mu_{\overline{w}}}\sum_{j':g(j')=\overline{w}} U_{j'}\mathcal{E}_{W\to A^n}(\rho_W)U_{j'}^\dagger\right)(U_jP) \tag{8.131}$$

$$= \widetilde{\mathcal{M}}_{A^n\to A^n}\mathcal{E}_{W\to A^n}(\rho_W). \tag{8.132}$$

Hence, it holds that $\overline{\mathcal{D}}_{B^n\to W}\circ\widetilde{\mathcal{M}}_{A^n\to A^n}\circ\mathcal{E}^{\overline{w}}_{W\to A^n} = \mathcal{D}_{B^n\to W}\circ\widetilde{\mathcal{M}}_{A^n\to A^n}\circ\overline{\mathcal{E}}_{W\to A^n} = c\mathrm{id}_W$ for $c \geqslant 1-\epsilon$. Finally for all ρ_W, we have

$$\|\rho^c_{B^n} - \rho^s_{B^n}\|_1 = \left\|\mathcal{M}^{\otimes n}_{A\to A}\circ\mathcal{E}_{W\to A^n}(\rho_W) - \frac{1}{\overline{M}}\sum_{\overline{w}}\overline{\mathcal{E}}^{\overline{w}}_{W\to A^n}(\rho_W)\right\|_1 \tag{8.133}$$

$$= \left\|\widetilde{\mathcal{M}}_{A^n\to A^n}\circ\mathcal{E}_{W\to A^n}(\rho_W) - \frac{1}{\overline{M}}\sum_{\overline{w}}\overline{\mathcal{E}}^{\overline{w}}_{W\to A^n}(\rho_W)\right\|_1 + \left\|\widetilde{\widetilde{\mathcal{M}}}_{A^n\to A^n}\circ\mathcal{E}_{W\to A^n}(\rho_W)\right\|_1. \tag{8.134}$$

For the first term, we have

$$\left\|\widetilde{\mathcal{M}}_{A^n\to A^n}\circ\mathcal{E}_{W\to A^n}(\rho_W) - \frac{1}{\overline{M}}\sum_{\overline{w}}\overline{\mathcal{E}}^{\overline{w}}_{W\to A^n}(\rho_W)\right\|_1 \tag{8.135}$$

$$= \left\|\sum_{j\in\mathcal{J}} F_j\mathcal{E}_{W\to A^n}(\rho_W)F_j^\dagger - \frac{1}{\overline{M}}\sum_{\overline{w}}\frac{1}{\mu_{\overline{w}}}\sum_{j:g(j)=\overline{w}} U_j\mathcal{E}_{W\to A^n}(\rho_W)U_j^\dagger\right\|_1 \tag{8.136}$$

$$= \left\|\sum_{j\in\mathcal{J}} F_j\mathcal{E}_{W\to A^n}(\rho_W)F_j^\dagger - \sum_{j\in\mathcal{J}} P_{g(J)}(j)U_j\mathcal{E}_{W\to A^n}(\rho_W)U_j^\dagger\right\|_1 \tag{8.137}$$

$$\overset{(a)}{=} \left\|\sum_{j\in\mathcal{J}} d_jU_j\mathcal{E}_{W\to A^n}(\rho_W)U_j^\dagger - \sum_{j\in\mathcal{J}} P_{g(J)}(j)U_j\mathcal{E}_{W\to A^n}(\rho_W)U_j^\dagger\right\|_1 \tag{8.138}$$

$$\leqslant \sum_{j\in\mathcal{J}}|d_j - P_{g(J)}(j)| \tag{8.139}$$

$$\leqslant \sum_{j\in\mathcal{J}} \left| d_j - \frac{d_j}{\sum_{j'\in\mathcal{J}} d_{j'}} \right| + \sum_{j\in\mathcal{J}} \left| \frac{d_j}{\sum_{j'\in\mathcal{J}} d_{j'}} - P_{g(J)}(j) \right| \leqslant \epsilon + \zeta, \quad (8.140)$$

where (a) follows since by the definition of $\mathcal{E}_{W\to A^n}$, $\Pi\mathcal{E}_{W\to A^n}(\rho_W)\Pi = \mathcal{E}_{W\to A^n}(\rho_W)$, and $F_j\Pi = \sqrt{d_j}U_j\Pi$. Furthermore,

$$\left\| \widetilde{\widetilde{\mathcal{M}}}_{A^n\to A^n} \circ \mathcal{E}_{W\to A^n}(\rho_W) \right\|_1 = \text{tr}\left(\widetilde{\widetilde{\mathcal{M}}}_{A^n\to A^n} \circ \mathcal{E}_{W\to A^n}(\rho_W) \right) \quad (8.141)$$

$$= \text{tr}\left(\mathcal{D}_{B^n\to W} \circ \widetilde{\widetilde{\mathcal{M}}}_{A^n\to A^n} \circ \mathcal{E}_{W\to A^n}(\rho_W) \right) \quad (8.142)$$

$$= \text{tr}\left(\mathcal{D}_{B^n\to W} \circ \left(\mathcal{M}_{A\to A}^{\otimes n} - \widetilde{\mathcal{M}}_{A^n\to A^n} \right) \circ \mathcal{E}_{W\to A^n}(\rho_W) \right) \quad (8.143)$$

$$= 1 - c \leqslant \epsilon. \quad (8.144)$$

□

8.8 Proof of Proposition 7

By the triangle inequality, we have

$$\left\| \sum_x P_X(x) \left(\mathcal{N}^x(\rho^x) - \sum_{x'} \mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}}) \right) \right\|_1 \quad (8.145)$$

$$\leqslant \left\| \sum_x P_X(x) \left(\mathcal{N}^x(\rho^x) - \mathcal{N}^x(\sqrt{\Lambda^x}\rho^x\sqrt{\Lambda^x}) \right) \right\|_1 + \left\| \sum_{x\neq x'} P_X(x)\mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}}) \right\|_1 . \quad (8.146)$$

Since $\mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}})$ is positive semi-definite, the second term would simplify as

$$\left\| \sum_{x\neq x'} P_X(x)\mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}}) \right\|_1 = \sum_{x\neq x'} P_X(x)\text{tr}\left(\mathcal{N}^{x'}(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}}) \right) \quad (8.147)$$

$$= \sum_{x\neq x'} P_X(x)\text{tr}\left(\sqrt{\Lambda^{x'}}\rho^x\sqrt{\Lambda^{x'}} \right) \quad (8.148)$$

$$= \sum_{x \neq x'} P_X(x) \text{tr}\left(\Lambda^{x'} \rho^x\right) \tag{8.149}$$

$$= 1 - \sum_x P_X(x) \text{tr}\left(\Lambda^x \rho^x\right) \leqslant \epsilon. \tag{8.150}$$

Furthermore, by the gentle measurement lemma, we have

$$\left\| \sum_x P_X(x) \left(\mathcal{N}^x(\rho^x) - \mathcal{N}^x(\sqrt{\Lambda^x}\rho^x\sqrt{\Lambda^x}) \right) \right\|_1 \tag{8.151}$$

$$\leqslant \sum_x P_X(x) \left\| \mathcal{N}^x(\rho^x) - \mathcal{N}^x(\sqrt{\Lambda^x}\rho^x\sqrt{\Lambda^x}) \right\|_1 \tag{8.152}$$

$$\overset{(a)}{\leqslant} \sum_x P_X(x) \left\| \rho^x - \sqrt{\Lambda^x}\rho^x\sqrt{\Lambda^x} \right\|_1 \tag{8.153}$$

$$\leqslant 2 \sum_x P_X(x) \sqrt{1 - \text{tr}\left(\Lambda^x \rho^x\right)} \tag{8.154}$$

$$\overset{(b)}{\leqslant} 2 \sqrt{1 - \sum_x P_X(x) \text{tr}\left(\Lambda^x \rho^x\right)} \leqslant 2\sqrt{\epsilon}, \tag{8.155}$$

where (a) follows from the date processing inequality (which holds for non-normalized states), and (b) follows from the concavity of the mapping $x \mapsto \sqrt{1-x}$.

8.9 Proof of Lemma 60

We extend $\{F_j\}_{j \in \mathcal{J}}$ to a Kraus representation $\{F_j\}_{j \in \mathcal{K}}$ for the channel $\mathcal{M}_{A \to A}$ with $\mathcal{J} \subset \mathcal{K}$. By [85], $U_{A \to AE} \triangleq \sum_{j \in \mathcal{K}} F_j \otimes |j\rangle_E$ is an isometric extension of $\mathcal{M}_{A \to A}$ where $\{|j\rangle_E\}_{j \in \mathcal{K}}$ is an orthonormal basis for the environment space $\mathcal{H}_E$. Let $\rho_A \triangleq \mathcal{E}_{W \to A}\left(\frac{1}{M}\mathbf{1}_W\right) = \frac{1}{M}\Pi$, and

$$\rho_E \triangleq \text{tr}_A(U_{A \to AE} \rho_A U_{A \to AE}^\dagger) \tag{8.156}$$

$$= \text{tr}_A \left(\left(\sum_{j \in \mathcal{K}} F_j \otimes |j\rangle_E \right) \left(\frac{1}{M}\Pi \right) \left(\sum_{j \in \mathcal{K}} F_j \otimes |j\rangle_E \right)^\dagger \right) \tag{8.157}$$

$$= \frac{1}{M} \sum_{j,j' \in \mathcal{K}} \text{tr}\left(F_j \Pi F_{j'}^\dagger\right) |j\rangle\langle j'|_E. \tag{8.158}$$

For the projector $\Gamma \triangleq \sum_{j\in\mathcal{J}} |j\rangle\langle j|_E$, we have

$$\begin{aligned}
\text{tr}\left(\Gamma\rho_E\right) &= \text{tr}\left(\left(\sum_{j\in\mathcal{J}} |j\rangle\langle j|_E\right)\left(\frac{1}{M}\sum_{j,j'\in\mathcal{K}} \text{tr}\left(F_j \Pi F_{j'}^\dagger\right)|j\rangle\langle j'|_E\right)\right) && (8.159)\\
&= \frac{1}{M}\sum_{j\in\mathcal{J}} \text{tr}\left(F_j\Pi F_j^\dagger\right) && (8.160)\\
&\overset{(a)}{=} \text{tr}\left(\widetilde{\mathcal{M}}_{A\to A}\left(\frac{1}{M}\Pi\right)\right) && (8.161)\\
&= \text{tr}\left(\widetilde{\mathcal{M}}_{A\to A}\left(\mathcal{E}_{W\to A}\left(\frac{1}{M}\mathbf{1}_W\right)\right)\right) && (8.162)\\
&\overset{(b)}{=} \text{tr}\left(\mathcal{D}_{A\to W}\left(\widetilde{\mathcal{M}}_{A\to A}\left(\mathcal{E}_{W\to A}\left(\frac{1}{M}\mathbf{1}_W\right)\right)\right)\right) && (8.163)\\
&= \text{tr}\left(c\text{id}_W\left(\frac{1}{M}\mathbf{1}_W\right)\right) && (8.164)\\
&= c \geqslant 1-\epsilon, && (8.165)
\end{aligned}$$

where (a) follows since $\{F_j\}_{j\in\mathcal{J}}$ is a Kraus representation of $\widetilde{\mathcal{M}}_{A\to A}$, (b) follows since $\mathcal{D}_{A\to W}$ is trace-preserving. By the gentle measurement lemma [123], we obtain that

$$\left\|\frac{\Gamma\rho_E\Gamma}{\text{tr}\left(\Gamma\rho_E\right)} - \rho_E\right\|_1 \leqslant 2\sqrt{\epsilon}. \tag{8.166}$$

We can write

$$\frac{\Gamma\rho_E\Gamma}{\text{tr}\left(\Gamma\rho_E\right)} = \frac{1}{\text{tr}\left(\Gamma\rho_E\right)}\left(\sum_{j\in\mathcal{J}}|j\rangle\langle j|_E\right)\left(\frac{1}{M}\sum_{j,j'\in\mathcal{K}}\text{tr}\left(F_j\Pi F_{j'}^\dagger\right)|j\rangle\langle j'|_E\right)\left(\sum_{j\in\mathcal{J}}|j\rangle\langle j|_E\right) \tag{8.167}$$

$$\begin{aligned}
&= \frac{1}{\text{tr}\left(\Gamma\rho_E\right)M}\sum_{j,j'\in\mathcal{J}}\text{tr}\left(F_j\Pi F_{j'}\dagger\right)|j\rangle\langle j'|_E && (8.168)\\
&= \frac{1}{\text{tr}\left(\Gamma\rho_E\right)M}\sum_j d_j|j\rangle\langle j|_E && (8.169)\\
&= \sum_{j\in\mathcal{J}} P_J(j)|j\rangle\langle j|_E. && (8.170)
\end{aligned}$$

Therefore, we have $H_{\min}^{\delta}(P_J) \geqslant H_{\min}^{\delta-2\sqrt{\epsilon}}(\rho_E)$

REFERENCES

[1] J. Gleick, *The information: a history, a theory, a flood*, 1st ed. Pantheon Books, 2011, ISBN: 9780375423727.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, 379–423, 1948.

[3] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[4] Y. Liang, H. V. Poor, S. Shamai, *et al.*, "Information theoretic security," *Foundations and Trends® in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[5] E. Jorswieck, S. Tomasin, and A. Sezgin, "Broadcasting into the uncertainty: Authentication and confidentiality by physical-layer processing," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1702–1724, Oct. 2015.

[6] A. O. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, 2003.

[7] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal of Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[8] V. Korzhik, G. Morales-Luna, and M. Lee, "On the existence of perfect stegosystems," in *Digital Watermarking*, ser. Lecture Notes in Computer Science, M. Barni, I. Cox, T. Kalker, and H.-J. Kim, Eds., vol. 3710, Springer Berlin Heidelberg, 2005, pp. 30–38, ISBN: 978-3-540-28768-1.

[9] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, vol. 192, no. 1, pp. 41–56, 2004.

[10] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. Morgan Kaufmann, 2007, Google-Books-ID: JZQLpzihtecC, ISBN: 9780080555805.

[11] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich, "The square root law of steganographic capacity," in *Proceedings of the 10th ACM workshop on Multimedia and security - MM&Sec '08*, ACM Press, 2008, p. 107, ISBN: 9781605580586.

[12] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[13] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[14] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, 752–772, 1993.

[15] K. S. K. Arumugam and M. R. Bloch, *Covert communication over a k-user multiple access channel*, accepted to *IEEE Transactions on Information Theory*, Jun. 2019. eprint: 1803.06007.

[16] K. S. K. Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2787–2801, Oct. 2019. eprint: 1808.09556.

[17] V. Y. F. Tan and S. Lee, "Time-Division is Optimal for Covert Communication Over Some Broadcast Channels," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1377–1389, May 2019.

[18] D. KIBLOFF, S. Perlaza, and L. Wang, "Embedding Covert Information on a Given Broadcast Code," in *ISIT 2019 - IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 1–5.

[19] A. Sheikholeslami, M. Ghaderi, D. Towsley, B. A. Bash, S. Guha, and D. Goeckel, "Multi-Hop Routing in Covert Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3656–3669, Jun. 2018.

[20] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[21] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Multilevel-coded pulse-position modulation for covert communications," in *2018 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2018, 1864–1868, ISBN: 9781538647813.

[22] Q. Zhang, M. Bakshi, and S. Jaggi, "Computationally efficient deniable communication," in *2016 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 2234–2238.

[23] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. of IEEE International Symposium on Information Theory*, Barcelona, Spain, 2016, pp. 2064–2068.

[24] J. M. Arrazola and V. Scarani, "Covert quantum communication," *Phys. Rev. Lett. 117, 250503 (2016)*, vol. 117, p. 250 503, 25 Apr. 19, 2016. arXiv: 1604.05438v3 [quant-ph].

[25] K. S. K. Arumugam and M. R. Bloch, "Keyless asynchronous covert communication," in *2016 IEEE Information Theory Workshop (ITW)*, Sep. 2016, pp. 191–195.

[26] B. A. Bash, D. Goeckel, and D. Towsley, "Covert Communication Gains From Adversary's Ignorance of Transmission Time," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8394–8405, Dec. 2016.

[27] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011, Google-Books-ID: Ack2AAAAQBAJ, ISBN: 9781139503143.

[28] A. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 21, no. 3, 294–300, 1975.

[29] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Transactions on Information Theory*, vol. 60, no. 11, 6760–6786, 2014.

[30] R. Renner, "Security of quantum key distribution," *International Journal of Quantum Information*, vol. 06, no. 01, 1–127, 2008.

[31] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2013, Google-Books-ID: dphmVbPSLHMC, ISBN: 9781107067844.

[32] M. A. Nielsen, I. L. Chuang, and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000, ISBN: 9780521635035.

[33] A. Uhlmann, "The transition probability in the state space of a *-algebra," *Reports on Mathematical Physics*, vol. 9, no. 2, pp. 273–279, Apr. 1976.

[34] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Reviews of Modern Physics*, vol. 81, no. 3, 1301–1350, 2009.

[35] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[36] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Phys. Rev. Lett.*, vol. 108, p. 200 501, 20 2012.

[37] M. Hayashi, *Quantum information theory: mathematical foundation*, Second edition, ser. Graduate texts in physics. Springer, 2017, ISBN: 9783662497258.

[38] M. Raginsky, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 1–2, 1–247, 2013.

[39] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, 2190–2212, 2019.

[40] ——, "Second-order asymptotics of covert communications over noisy channels," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, 2224–2228.

[41] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory*, Istanbul, Turkey, 2013, pp. 2945–2949.

[42] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. of IEEE Information Theory Workshop*, Cambridge, UK, 2016, pp. 364–368.

[43] K. S. K. Arumugam and M. R. Bloch, "Keyless covert communication over multiple-access channels," in *Proc. of IEEE International Symposium on Information Theory*, Barcelona, Spain, 2016, pp. 2229–2233.

[44] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. of IEEE International Symposium on Information Theory*, Honolulu, HI, 2014, pp. 601–605.

[45] M. Tahmasbi and M. R. Bloch, "Second-order asymptotics of covert communications over noisy channels," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2224–2228.

[46] M. R. Bloch and S. Guha, "Optimal covert communications using pulse-position modulation," in *Proc. of IEEE International Symposium on Information Theory*, Aachen, Germany, Jun. 2017, pp. 2835–2839.

[47] W. Yang, R. F. Schaefer, and H. V. Poor, "Finite-blocklength bounds for wiretap channels," in *Proc. IEEE International Symposium on Information Theory*, Barcelona, Spain, 2016, pp. 3087–3091.

[48] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1–2, 1–184, 2014.

[49] M. Tahmasbi, A. Savard, and M. R. Bloch, "Covert capacity of non-coherent rayleigh-fading channels," *arXiv:1810.07687 [cs, math]*, 2018, arXiv: 1810.07687.

[50] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communications on continuous-time channels in the presence of jamming," in *Proc. of 51st Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2017, pp. 1697–1701.

[51] B. A. Bash, D. Goeckel, and D. Towsley, "Covert communication gains from adversary's ignorance of transmission time," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8394–8405, Dec. 2016.

[52] S. Yan, B. He, X. Zhou, Y. Cong, and A. L. Swindlehurst, "Delay-intolerant covert communications with either fixed or random transmit power," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 129–140, Jan. 2019.

[53] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4766–4779, Jul. 2018.

[54] J. G. Smith, "On the information capacity of peak and average power constrained gaussian channels," PhD thesis, University of California, Berkeley, 1969.

[55] ——, "The information capacity of amplitude and variance-constrained scalar gaussian channels," *Information and Control*, vol. 18, pp. 203–219, 1971.

[56] I. Abou-Faycal, M. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, 1290–1301, 2001.

[57] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[58] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd. Wiley-Interscience, 2006.

[59] J. Hou, "Coding for relay networks and effective secrecy for wire-tap channels," PhD thesis, 2014.

[60] M. J. Ablowitz and A. S. Fokas, *Complex variables: introduction and applications*. Cambridge University Press, 2003.

[61] R. B. Ash, *Information Theory*, ser. Interscience Tracts in Pure and Applied Mathematics. John Wiley & Sons, 1965.

[62] D. G. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1997.

[63] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2009.

[64] S. Y. Park and A. K. Bera, "Maximum entropy autoregressive conditional heteroskedasticity model," *Journal of Econometrics*, vol. 150, no. 2, pp. 219–230, 2009.

[65] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2010.

[66] M. Tahmasbi and M. R. Bloch, *Covert communication when the code is unknown at the warden*, accepted to *Allerton Conference*, Aug. 2019.

[67] S. Watanabe and M. Hayashi, "Strong converse and second-order asymptotics of channel resolvability," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 1882–1886.

[68] M. Tahmasbi and M. R. Bloch, "Covert secret key generation with an active warden," *IEEE Transactions on Information Forensics and Security*, vol. 15, 1026–1039, 2020.

[69] ——, "Covert secret key generation," in *2017 IEEE Conference on Communications and Network Security (CNS)*, IEEE, 2017, 540–544, ISBN: 9781538606834.

[70] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Info. Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[71] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Info. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[72] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

[73] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, "Information-theoretically secret key generation for fading wireless channels," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 240–254, 2010.

[74] P. Lin, C. R. Janda, and E. A. Jorswieck, "Stealthy secret key generation," in *Proc. of IEEE Global Conference on Signal and Information Processing*, 2017, pp. 492–496.

[75] M. Tahmasbi, M. R. Bloch, and A. Yener, "Learning an adversary's actions for secret communication," *IEEE Transactions on Information Theory*, 1–1, 2019.

[76] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd. Cambridge University Press, 2011.

[77] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Transactions on Information Theory*, vol. 62, no. 4, 1836–1849, 2016.

[78] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.

[79] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, Mar. 1963.

[80] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[81] M. Tahmasbi and M. R. Bloch, "Framework for covert and secret key expansion over classical-quantum channels," *Physical Review A*, vol. 99, no. 5, p. 052 329, 2019.

[82] J. M. Arrazola and R. Amiri, "Secret-key expansion from covert communication," *Phys. Rev. A*, vol. 97, p. 022 325, 2 2018. arXiv: 1708.09103v1 [quant-ph].

[83] Y. Liu, J. M. Arrazola, W.-Z. Liu, W. Zhang, I. W. Primaatmaja, H. Li, L. You, Z. Wang, V. Scarani, Q. Zhang, and J.-W. Pan, *Experimental unconditionally secure covert communication in dense wavelength-division multiplexing networks*, Sep. 20, 2017. arXiv: 1709.06755v1 [quant-ph].

[84] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, pp. –, 2015.

[85] M. Wilde, *Quantum information theory*. Cambridge University Press, 2013, ISBN: 9781107034259.

[86] M. Tahmasbi and M. R. Bloch, "Toward undetectable quantum key distribution over bosonic channels," *arXiv:1904.12363 [quant-ph]*, 2019, arXiv: 1904.12363.

[87] S. Watanabe and M. Hayashi, "Non-asymptotic analysis of privacy amplification via rényi entropy and inf-spectral entropy," in *2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, 2013, pp. 2715–2719.

[88] M. Tomamichel, *Quantum information processing with finite resources: mathematical foundations*, ser. Springer briefs in mathematical physics. Springer, 2016, ISBN: 9783319218908.

[89] J. Martinez-Mateo, D. Elkouss, and V. Martin, "Key Reconciliation for High Performance Quantum Key Distribution," *Scientific Reports*, vol. 3, p. 1576, 2013.

[90] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: A random binning analogy," *IEEE Trans. Info. Theory*, vol. 62, no. 5, pp. 2410–2429, 2016.

[91] M. Tahmasbi and M. R. Bloch, "A framework for covert and secret key expansion over quantum channels," *arXiv preprint arXiv:1811.05626*, 2018.

[92] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner, "Continuous variable quantum key distribution: Finite-key analysis of composable security against coherent attacks," *Phys. Rev. Lett.*, vol. 109, p. 100 502, 10 2012.

[93] A. Leverrier, R. García-Patrón, R. Renner, and N. J. Cerf, "Security of continuous-variable quantum key distribution against general attacks," *Phys. Rev. Lett.*, vol. 110, p. 030 502, 3 2013.

[94] R. Renner and J. I. Cirac, "De finetti representation theorem for infinite-dimensional quantum systems and applications to quantum cryptography," *Phys. Rev. Lett.*, vol. 102, p. 110 504, 11 2009.

[95] Q. Zhuang, Z. Zhang, J. Dove, F. N. C. Wong, and J. H. Shapiro, "Floodlight quantum key distribution: A practical route to gigabit-per-second secret-key rates," *Physical Review A*, vol. 94, no. 1, p. 012 322, 2016.

[96] I. Sason and S. Verdú, "F -divergence inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, 5973–6006, 2016.

[97] M. Bellare and S. Tessaro, "Polynomial-time, semantically-secure encryption achieving the secrecy capacity," *arXiv preprint arXiv:1201.3160*, 2012.

[98] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, 3989–4001, 2011.

[99] M. Tahmasbi and M. Bloch, "Steganography protocols for quantum channels," *arXiv:1907.09602 [quant-ph]*, 2019, arXiv: 1907.09602.

[100] M. Tahmasbi and M. R. Bloch, "Steganography protocols for quantum channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, 2179–2183.

[101] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 563–593, 2003.

[102] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2706–2722, 2008.

[103] A. D. Ker, "A capacity result for batch steganography," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 525–528, 2007.

[104] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha, "Covert wireless communication with artificial noise generation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7252–7267, 2018.

[105] K. Shahzad, X. Zhou, S. Yan, J. Hu, F. Shu, and J. Li, "Achieving covert wireless communications using a full-duplex receiver," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8517–8530, 2018.

[106] S. Natori, "Why quantum steganography can be stronger than classical steganography," in *Quantum Computation and Information: From Theory to Experiment*, H. Imai and M. Hayashi, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 235–240, ISBN: 978-3-540-33133-9.

[107] J. Gea-Banacloche, "Hiding messages in quantum data," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4531–4536, Aug. 2002.

[108] B. A. Shaw and T. A. Brun, "Quantum steganography with noisy quantum channels," *Phys. Rev. A*, vol. 83, p. 022 310, 2 2011.

[109] C. Sutherland and T. A. Brun, *Quantum steganography over noisy channels: Achievability and bounds*, arXiv preprint 1808.03183, Aug. 2018.

[110] B. Sanguinetti, G. Traverso, J. Lavoie, A. Martin, and H. Zbinden, "Perfectly secure steganography: Hiding information in the quantum noise of a photograph," *Phys. Rev. A*, vol. 93, p. 012 336, 1 2016.

[111] C. Sutherland and T. A. Brun, *Quantum steganography over noiseless channels: Achievability and bounds*, arXiv: 1805.01599, May 2018.

[112] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7377–7385, 2011.

[113] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7693–7710, 2013.

[114] I. Devetak, A. W. Harrow, and A. J. Winter, "A resource framework for quantum shannon theory," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4587–4618, 2008.

[115] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A*, vol. 53, pp. 2046–2052, 4 1996.

[116] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 1175–1187, 2013.

[117] G. Adesso, S. Ragy, and A. R. Lee, "Continuous variable quantum information: Gaussian states and beyond," *Open Systems & Information Dynamics*, vol. 21, no. 01n02, p. 1 440 001, 2014. eprint: `https://doi.org/10.1142/S1230161214400010`.

[118] R. Klesse, "Approximate quantum error correction, random codes, and quantum channel capacity," *Phys. Rev. A*, vol. 75, p. 062 315, 6 2007.

[119] P. Hayden, P. W. Shor, and A. Winter, "Random quantum codes from gaussian ensembles and an uncertainty relation," *Open Systems & Information Dynamics*, vol. 15, no. 01, pp. 71–89, 2008.

[120] P. Hayden and A. Winter, "Counterexamples to the maximal p-norm multiplicativity conjecture for all p ¿ 1," *Communications in Mathematical Physics*, vol. 284, no. 1, pp. 263–280, 2008.

[121] T. Ogawa and M. Hayashi, "On error exponents in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1368–1372, 2004.

[122] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5840–5847, 2009.

[123] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481–2485, 1999.